Searching for the Source

Using Shallow Groundwater Chemistry to Determine the Sources for Streamflow in a Pre-Alpine Catchment

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde (Dr. sc. nat)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Leonie Adriana Kiewiet

aus

den Niederlanden

Promotionskommission

Prof. Dr. Jan Seibert (Vorsitz)
Dr. Ilja van Meerveld (Leitung der Dissertation)
Prof. Dr. Markus Egli
Dr. Manfred Stähli

Zürich, 2020

Everywhere I go, you're always right beside me.
you're more reliable than rain or snow, and never cease to find me.

You warm me when I'm cold, and cool me when it's hot,
without you I would freeze or fade, but with your back I'm not.

And even if others insist, and keep me occupied,
you wait until the time is right, so I won't grow too wide.

I'm relying on your faithfulness, for most I'll ever be,
until we merge and flow as one, united to the sea.

Contents

List of Figures

List of Tables

Publications and contributions

The journal articles included in this Ph.D. dissertation are listed below. They are added to the dissertation as appendices, and further referred to in the text with roman letters.

I **L. Kiewiet**, J. von Freyberg, H.J. van Meerveld, "Spatiotemporal variability in hydrochemistry of shallow groundwater in a small pre-alpine catchment: The importance of landscape elements", Hydrol. Process., 33, 2502– 2522, 2019.

Ilja van Meerveld and I designed the study. I collected the samples and field data, completed the corresponding laboratory analyses, did the data analyses, and wrote the first draft of the article. Ilja van Meerveld provided feedback for the data collection and analyses. All co-authors were active in discussions of results and contributed to the writing process.

II **L. Kiewiet**, H.J. van Meerveld, M. Stähli, J. Seibert, "Do stream water solute concentrations reflect when connectivity occurs in a small, pre-Alpine headwater catchment?", Hydrol. Earth Syst. Sci., 24, 3381–3398, 2020. .

I defined the sampling strategy with help from Ilja van Meerveld, and collected and analysed the samples and field data for the study, including laboratory analyses. Michael Rinderer simulated the extent

of the hydrologically connected area. All co-authors contributed to discussions of the results. I wrote the first draft of the article; all co-authors contributed to the writing process.

III **L. Kiewiet**, H.J. van Meerveld, J. Seibert "Effects of spatial variability in the groundwater isotopic composition in a pre-alpine headwater catchment on hydrograph separation results", Water Resour. Res., 56, e2019WR026855, 2020.

All co-authors were involved in the design of the study. I collected and analysed the samples and field data for this study, and wrote the first draft of the article. All co-authors contributed to the interpretation and discussion of the results, and contributed to the writing process.

Summary

Streamflow in undisturbed catchments in humid temperate climates consists predominantly of groundwater, even during large rainfall events. Groundwater contributions to streamflow sustain stream baseflow, influence stream water temperature, and affect stream water chemistry. Despite the obvious importance of groundwater for streamflow and water management, only few studies have focused on the spatial variability in in its isotopic and chemical composition. Although we might expect that groundwater chemistry differs from location to location due to differences in subsurface pathways and residence times, groundwater chemistry is often assumed to be relatively uniform, and to be well-represented by stream baseflow.

There are at least two problems with this status quo. First, we miss out on learning about flow pathways within catchments, and which areas contribute to streamflow during different stages of wetness. Second, assumptions of homogeneity can be problematic for source-area analyses if the spatial variability leads to temporal variability in the composition of the groundwater that actually contributes to streamflow due to the expansion of the hydrologically connected and contributing area during rainfall events. If the composition of transiently connected groundwater sources is significantly different from the permanently connected groundwater, changes in stormflow chemistry might be (partially) caused by contributions from previously disconnected ground-

water sources, rather than the inferred contributions from soil water or rain water. This can lead to large uncertainties in the results of source area analyses, and might go unnoticed if we are not aware of the spatial variability in the groundwater chemistry and isotopic composition.

Therefore, this thesis aims to enhance our knowledge of the spatial variability in shallow groundwater composition in a small pre-alpine headwater catchment by quantifying the variability and investigating which processes cause this variability. This information was then used to identify the source areas for baseflow and stormflow, and to estimate the uncertainty in hydrograph separation analyses due to the spatial variability in the groundwater composition. The analyses are based on spatial sampling of groundwater, soil water and streamwater during nine baseflow snapshot campaigns, and additional sampling of streamwater and rainwater during four rainfall events in the pre-alpine Studibach catchment (Alptal), using a measurement network of 51 groundwater wells, 18 suction lysimeters, seven stream gauges and three rain gauges.

The results of the groundwater sampling showed that the spatial variability in shallow groundwater chemistry was large, and for most parameters larger than the temporal variability. Differences from the catchment average concentrations were used to distinguish four shallow groundwater types, of which three are related to landscape elements: Type I: riparian-like areas, characterized by high concentrations of iron, manganese and cobalt; Type II: hillslopes, characterized by high concentrations of copper, zinc and nickel; and Type III: 'deep' groundwater, which had a depleted isotopic signature compared to the other types, and high concentrations of strontium. Groundwater Type IV was influenced by bedrock with a different composition, and had high concentrations of sulfate and magnesium. The groundwater compositions and types were only weakly related to topographic and hydrodynamic site attributes. However, soil and bedrock leachates, and subsurface topographic data (obtained from geophysical profiles) confirmed the spatial distribution and chemical composition characteristics for each groundwater type.

Baseflow was not an equal mixture of the different groundwater types. For the majority of the sampling campaigns baseflow chemistry most strongly resembled riparian-like groundwater (I) for all but one sub-catchment. However, similarity to the hillslope-type groundwater (II) was larger shortly after snowmelt, reflecting differences in hydrologic connectivity. Three-component end-member mixing analyses indicated that groundwater dominated stormflow, and that soil water fractions were minimal for three of the four events. However, the large variability in the soil and groundwater composition compared to the changes in stormflow composition led to large uncertainties. Stormflow was not a conservative mixture of rainwater and baseflow, which indicates that other sources (soil water or other groundwater sources) induced changes in streamflow composition during events. Streamwater chemistry changed gradually during events, which likely reflects a gradual increase in the hydrologically connected and contributing area rather than a threshold type behaviour.

The hydrograph separation results depended highly on the choice of which and how many samples were used to characterize the pre-event water composition. Generally, including more samples yielded more robust results. The number of groundwater samples needed to characterize the average groundwater composition and its variability was much larger than is common in small-scale hydrograph separation studies. However, by taking a few more samples than is typical, one can already obtain an estimate of the variability. Analyses that do not include the variability in pre-event water composition (for instance by taking only a baseflow sample) likely underestimate the real uncertainty.

Overall, this thesis demonstrates that an improved understanding and representation of shallow groundwater chemistry is useful for studies in small headwater catchments. Therefore, it is recommended that the hydrologic community uses more information on the variability in the groundwater composition and subsurface topography in hydrological analyses. Increasing the number of groundwater samples is important for robust analyses. This should inspire us to go out and measure. This mes-

sage is especially important in the age of declining field research and for climatic regions for which we currently have limited data. Lastly, the results of fundamental analyses, like hydrograph separation, shape our conceptual understanding, and thereby influence the way we develop future research or models to investigate hydrological processes at larger scales. As such, getting the fundamentals about groundwater and its variability right is paramount.

Zusammenfassung für Laien

In naturbelassenen Gebieten der gemässigten Klimazone besteht der Abfluss in Bächen und Flüssen hauptsächlich aus Grundwasser. Damit prägt das Grundwasser den Abfluss sowohl in seiner Menge als auch in seiner Temperatur und chemischen Zusammensetzung. Trotz der grossen Bedeutung des Grundwassers für Abfluss und Wassermanagement, wissen wir relativ wenig über die räumliche Variabilität seiner chemischen Zusammensetzung. Diese Zusammensetzung hängt davon ab, entlang welcher Fliesswege das Wasser fliesst und wie lange es schon unter der Erde ist, und kann daher räumlich varieren. Die räumliche Variation ist jedoch wenig untersucht, und daher wird häufig angenommen, dass die Grundwasserzusammensetzung überall in etwa gleich ist und während trockenem Wetter der Zusammensetzung des Abflusses gleicht.

Diese Annahme ist aus zwei Gründen problematisch. Erstens fehlt uns damit die Kenntnis darüber, wo das Grundwasser fliesst und gespeichert wird. Dieses Wissen könnte beispielsweise dazu genutzt werden, um Entscheidungen über die Verteilung von Wasserressourcen zu treffen oder um abzuschätzen, wie lange es dauert, bis kontaminiertes Grundwasser sich ausbreitet. Zweitens kann die zuvor beschriebene Annahme ein Problem für Analysen sein, bei denen wir untersuchen, woher der Abfluss kommt. Die Ergebnisse solcher Quellflächenanalysen können in hydrologischen Modellen zur Vorhersage von Hoch- und Niedrigwasser verwendet werden. Dabei wird häufig die chemische Zusammensetzung

von Abfluss, Bodenwasser, Grundwasser und Regenwasser verglichen und der Beitrag jeder Komponente zur Wasserzusammensetzung berechnet. Hierfür ist es jedoch wichtig, die genaue Zusammensetzung jeder Komponente zu kennen.

Deshalb habe ich die räumliche Variabilität der chemischen Zusammensetzung des Grundwassers in einem kleinen Gebiet in den Schweizer Voralpen, dem Alptal, untersucht. Starke Niederschläge führen hier häufig oft zu Überschwemmungen. Daher werden im Alptal schon seit längerem hydrologische Untersuchungen durchgeführt, und eine Ergänzung der bisherigen Messreihen mit den oben genannten Grundwassermessungen hat sich angeboten. Um die räumliche Variabilität der Grundwasserzusammensetzung zu bestimmen, verwendete ich 51 Grundwasserrohre in einem 20 Hektar grossen Einzugsgebiet. Ich habe das Grundwasser an diesen Rohren über einen Zeitraum von zwei Jahren neunmal beprobt. Während dieser Kampagnen, die ausschliesslich bei trockenem Wetter durchgeführt wurden, nahm ich am selben Tag auch Abfluss- und Bodenwasserproben. Zusätzlich habe ich bei vier Regenereignissen Proben eingeholt. Die chemische Zusammensetzung aller Proben habe ich im Labor bestimmt, und diese Daten habe ich verwendet, um den Beitrag jeder Komponente zur Wasserzusammensetzung zu berechnen. Um mögliche Ursachen der lokalen Variabilität in der Grundwasserzusammensetzung zu untersuchen, habe ich die Zusammensetzung des Grundwassers mit der Topographie und dem Untergrund verglichen, der mittels geophysikalischer Profile untersucht wurde.

Anhand der Daten habe ich festgestellt, dass die räumliche Variabilität der chemischen Zusammensetzung im Grundwasser gross ist; sie ist sogar grösser als die jahreszeitliche Veränderung. Die Grundwasserrohre im Testgebiet können in vier verschiedene Grundwassertypen unterteilt werden: Typ I, welcher Grundwasser in Gebieten mit Flussnähe oder in relativ flachen Gefilden umfasst; Typ II, welcher Grundwasser an Hängen oder an Bergrücken beschreibt; Typ III, welcher tiefes Grundwasser bezeichnet, und Typ IV, bei welchem das Grundwasser hauptsächlich

durch den Kontakt mit einem Grundgestein beeinflusst wurde, welches sich vom restlichen Grundgestein des Einzugsgebiets unterscheidet. Die chemischen Konzentrationen im Grundwasser waren nur schwach mit der Topographie korreliert. Die geophysikalischen Profile hingegen zeigten eine Verbindung von Hangneigung und Bodentiefe mit den Grundwassertypen.

Der Abfluss war vor allem eine Mischung aus Grundwassertyp I und II. In den Messkampagnen kurz nach der Schneeschmelze (Mai und Juni) ähnelte das Flusswasser dem Grundwasser des Typs II, später im Jahr des Typs I. Dies zeigt, dass der Beitrag der unterschiedlichen Komponenten zur Zusammensetzung des Flusswassers davon abhängt, wie nass das Gebiet ist. Das Flusswasser glich auch während der Regenereignisse grösstenteils dem Grundwasser, und weniger dem Regenwasser oder dem Bodenwasser. Bei diesen Wasserproben war die Unsicherheit, die abhängig ist von die räumliche Variabilität der chemischen Zusammensetzung, aber zu gross, um zu bestimmen, welcher Grundwassertyp dominierte. Es war jedoch deutlich zu sehen, dass bei drei der vier Regenereignisse das Bodenwasser den Afluss kaum beeinflusst hatte.

Mit meiner Arbeit konnte ich nachweisen, dass wir dank mehr Informationen über die räumliche Variabilität des Grundwassers untersuchen können, unter welchen Voraussetzungen - nass oder trocken - welche Art von Grundwasser die Zusammensetzung des Abflusses massgeblich bestimmt. Darüberhinaus habe ich gezeigt, dass die Unsicherheit für Komponentenanalysen gross sein kann. Dies kann zu falschen Interpretationen führen, wenn die Variabilität des Grundwassers nicht ausreichend berücksichtigt wird. Daraus können wir schliessen, dass es immer noch sehr wichtig ist, im Gelände Proben zu sammeln, auch wenn heutzutage zunehmend mit Computern und Modellen geforscht wird. Messdaten von Wasserproben aus dem Gelände werden insbesondere dazu benötigt, um Modelle zu testen und zu kalibrieren, damit wir sicherstellen können, dass wir unsere Entscheidungen auf einer robusten Analyse basieren.

Samenvatting voor familie en vrienden

In gebieden met een gematigd klimaat die niet extreem aangetast zijn door de mens komt het rivierwater merendeels van grondwater. Grondwater beïnvloedt niet alleen de rivierafvoer, maar ook de temperatuur en de chemische samenstelling van rivierwater. Ondanks het belang van grondwater voor afvoer- en watermanagement weten we vrij weinig over de ruimtelijke variabiliteit van de chemische samenstelling van grondwater. Die samenstelling hangt af van hoelang het water al ondergronds is en waar het gestroomd heeft, en kan dus verschillend zijn op verschillende plekken. Omdat hier niet veel informatie over is, wordt vaak aangenomen dat het grondwater overal ongeveer hetzelfde, of dat de chemische samenstelling van rivierwater in droge periodes (d.w.z., zonder directe invloed van regen) representatief is voor het grondwater.

Er zijn twee problemen met deze aannames. Ten eerste missen we een stuk informatie waarmee we mogelijk meer kunnen leren over hoe het water ondergronds stroomt en opgeslagen is. Deze informatie is belangrijk omdat we als hydrologen juist meer willen weten over de waterhuishouding in verschillende gebieden. Deze kennis kan bijvoorbeeld gebruikt worden om beslissingen te maken over het verdelen van water, of om te weten hoe snel een vervuiling zich verspreidt in het grondwater. Ten tweede kan het een probleem zijn voor analyses waarin we onderzoeken waar het rivierwater vandaan komt. De resultaten van zulke "bronanalyses" kunnen worden gebruikt in hydrologische modellen voor

het voorspellen van (extreem) hoog of laag water, of voor toekomstprojecties. Bronanalyses worden vaak gedaan door de chemische samenstelling van rivierwater, regen, grondwater en bodemwater te vergelijken, en de bijdrage van iedere bron te berekenen. Dan is het uiteraard belangrijk om de samenstelling van elke bron te weten.

Tijdens mijn doctoraat heb ik de ruimtelijke variabiliteit van de chemische samenstelling van grondwater in een klein gebied in de Zwitserse Voor-Alpen (Studibach, Alptal) onderzocht. Dit gebied is karakteristiek voor de regio. Regenval gaat hier vaak gepaard met wateroverlast. Daarom is er al sinds lange tijd hydrologisch onderzoek in dit gebied en was het ook voor dit onderzoek een geschikte locatie. Om de ruimtelijke variabiliteit van de chemische samenstelling van het ondiepe grondwater te bepalen, heb ik gebruik gemaakt van 51 peilbuizen in een stroomgebied van 20 hectare. Over een periode van twee jaar heb ik de peilbuizen negen keer bemonsterd. Tijdens deze campagnes, die uitsluitend tijdens droog weer uitgevoerd werden, heb ik ook monsters van het rivierwater en bodemwater genomen. Daarnaast heb ik tijdens vier regenbuien ook monsters van de regen en het rivierwater verzameld. Van alle monsters heb ik in het laboratorium de chemische samenstelling bepaald, en deze informatie heb ik gebruikt om de bijdrage van iedere bron aan het rivierwater te berekenen. De samenstelling van het grondwater heb ik ook vergeleken met de topografie en de ondergrond (d.m.v. geofysische profielen), om te onderzoeken wat de oorzaak kan zijn van verschillen in de grondwatersamenstelling op verschillende locaties.

Deze data tonen aan dat de ruimtelijke variabiliteit groot is, zelfs groter dan de seizoenale veranderingen. Aan de hand van de chemische samenstelling kunnen de peilbuizen in het gebied opgedeeld worden in vier verschillende grondwatertypes: Type I, dicht bij de oever of in vlakke gebieden; Type II, op hellingen; Type III, diep grondwater; en Type IV, beïnvloed door contact met een gesteente met een andere compositie. De concentraties van verschillende elementen (o.a. calcium, magnesium, ijzer, koper) in het grondwater waren niet sterk gecorreleerd met de topografie, terwijl de geofysische profielen lieten zien dat het

grondwatertype samenhangt met de combinatie van helling aan het oppervlak en bodemdiepte.

Het rivierwater was nooit een gelijke mix van alle verschillende grondwater types, maar vooral een mix van Type I en Type II. In de campagnes in mei en juni, direct na de sneeuwsmelt periode, leek de samenstelling van het rivierwater veel op Type II, terwijl het later in het jaar meer op Type I leek. Dit laat zien dat de bijdrage van verschillende grondwatertypes afhankelijk is van hoe nat het gebied is. Ook tijdens de regenbuien bestond het rivierwater grotendeels uit grondwater.

Het laatste punt dat ik onderzocht heb was de onzekerheid in bronanalyses als we de ruimtelijke variabiliteit niet hebben bepaald. Dit is het geval als we bijvoorbeeld maar één of een paar grondwater monsters hebben genomen. Uit deze analyse werd duidelijk dat de resultaten van de bronanalyses sterk afhankelijk zijn van hoeveel en welke monsters gebruikt worden in de analyse, en dat het nemen van meer monsters de onzekerheid verkleind. Voor de Studibach was de vermindering van de onzekerheid het sterkst voor de analyse met zes in plaats van drie grondwater monsters. Bovendien, na het nemen van meerdere monsters (d.w.z. meer dan één of twee) weet men waarschijnlijk al of er een grote variabiliteit is of niet, en had de onzekerheid dus beter ingeschat kunnen worden.

Met dit proefschrift laat ik zien dat we met informatie over de samenstelling van grondwater op verschillende locaties kunnen onderzoeken welk grondwater bijdraagt aan rivierwater en hoe dit verandert tijdens natte of droge periodes, en dat de onzekerheid voor dit soort analyses groter is als we de variabiliteit niet goed karakteriseren. Wat we hieruit kunnen concluderen is dat het waardevol is om naar buiten te gaan en monsters te verzamelen. Dit is een belangrijke boodschap omdat onderzoek tegenwoordig steeds meer met computers en modellen gedaan wordt. Juist om onze modellen te testen en te kalibreren hebben we data nodig, zodat we er zeker van kunnen zijn dat we onze beslissingen baseren op een robuuste analyse.

Chapter 1

Introduction

Streamflow in undisturbed catchments in humid temperate climates consists predominantly of groundwater, even during large rainfall events (Buttle, 1994; Klaus and McDonnell, 2013; Laudon and Slaymaker, 1997). Groundwater sustains baseflow (Kendall and Mcdonnell, 1998; Soulsby et al., 1998; Tetzlaff and Soulsby, 2008), and affects stream water quality (Soulsby et al., 2006) because it provides an influx of chemical compounds that were acquired along the subsurface flow pathway. Groundwater also affects stream temperature (Loheide and Gorelick, 2006) and provides thermal refugia (Brunke and Gonser, 1997). Despite the importance of groundwater for streamflow, little is known about the spatial variability in its chemical composition in small headwater catchments, where groundwater flows into headwater streams, and how this is affected by topography.

The chemical composition of groundwater at a certain location is related to the flow pathways, inflows from uphill and the surface, mixing, interactions with the soil and bedrock, and variability in flushing frequency. Consequently, different parts of the catchment may have a different groundwater composition. Previous research has shown that the isotopic composition and electrical conductivity (EC) (Kendall et al., 2001; Penna et al., 2016), base cation concentrations (Burns et al., 1998) and nitrate concentrations (Welsch et al., 2001) can differ markedly in shallow groundwater wells that are located on small distances (i.e., a few meters). Kendall

et al. (2001) reported a large variability in the isotopic composition of all subsurface water components for the artificial Hydrohill catchment, and attributed this variability to differences in flow conditions, such as macropore flow or matrix flow. Burns et al. (1998) found that the variability in base cation concentrations was inversely related to the accumulated area of the bedrock topography and differences in flushing frequency. Welsch et al. (2001) hypothesized that the variability in nitrate concentrations were caused by differences in flushing frequency between upslope and downslope locations.

Previous studies have shown that there is a large variability in groundwater level dynamics across hillslopes and headwater catchments (Bachmair and Weiler, 2012; Tromp-van Meerveld and McDonnell, 2006) and that these differences are related to topographically defined landforms (Detty and McGuire, 2010; Rinderer et al., 2015), or vary as a function of distance to the stream (Haught and van Meerveld, 2011; Penna et al., 2016; Seibert et al., 2011). These studies have also shown that groundwater levels in lower hillslope, or riparian zones, are more strongly correlated with streamflow than groundwater levels in upper hillslope zones (Detty and McGuire, 2010; Haught and van Meerveld, 2011; Seibert et al., 2003). Hence, if we expect that groundwater composition depends on the composition of the aquifer and local water fluxes, and that local water fluxes are determined by topography, we might hypothesize that the groundwater composition is also related to topography.

Hydrologic connectivity, or "the linkage of separate regions of a catchment via water flow" (Blume and van Meerveld, 2015), is also modulated by topography (Figure 1.1). While the source areas to the stream are usually limited to areas close to the stream during baseflow conditions, locations farther from the stream can contribute to streamflow during periods of high flow (Aubert et al., 2013; Blume and van Meerveld, 2015; Bracken and Croke, 2007; Jencso and McGlynn, 2011; Stieglitz et al., 2003). The expansion of the hydrologically connected area occurs during events and/or seasonally (Detty and McGuire, 2010; Jencso et al., 2009; van Meerveld et al., 2015). Hydrologic connectivity is dictated by the distribution of landscape elements and overall catchment wetness, and is one

of the major controls on the changes in stream water composition during events (Blume and van Meerveld, 2015; Bracken and Croke, 2007; Stieglitz et al., 2003). During periods of higher connectivity, (ground)water from different areas, and possibly different chemical signatures, can enter the stream (Martin et al., 2004; Seibert et al., 2009).

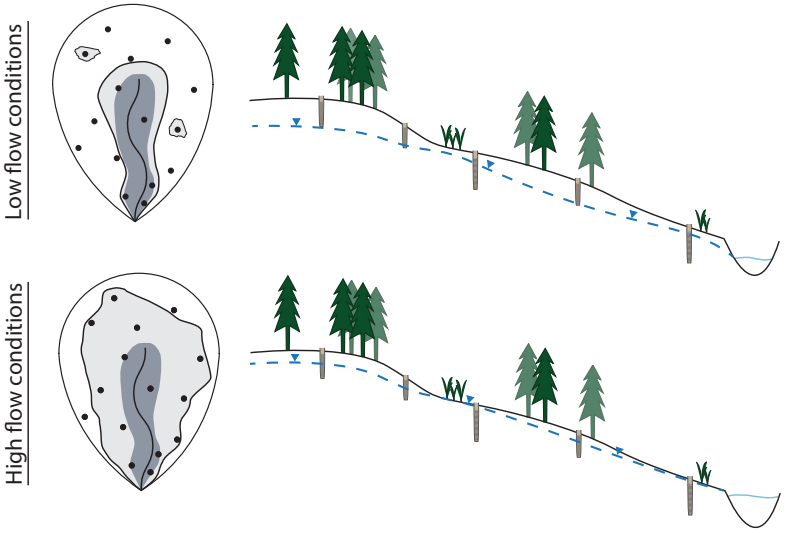

Figure 1.1 – Schematic drawing showing the plan view (left) and profile view (right) of the extent of the hydrologically connected area during low flow (top) and high flow (bottom) conditions. In plan view, the shading indicates if this source area is permanently connected (dark grey), permanently disconnected (white) or transiently connected (light grey); groundwater wells are indicated as black dots. The profiles show the groundwater table (blue dashed lines) multiple wells (light brown), and its relation to stream stage height (light blue line).

Studying the gradual or instantaneous changes in solute concentrations or isotopic composition of streamwater can thus help us to better understand how the contributions from different source areas change during rainfall events and when certain groundwater source areas contribute to stream-flow (Birkel and Soulsby, 2015; Bonell, 1998). Alternatively, changes in solute concentrations can be indicative for reactive transport and serve as

an indicator for biogeochemical processes. For instance, changes in the concentration of dissolved organic carbon (DOC) can reflect the inflow from areas with high organic carbon production, whereas increasing concentrations of redox-sensitive elements, such as iron, can reflect the contributions from source waters with a different oxidation state (Kaushal et al., 2018).

One can also use multiple tracers simultaneously in source-area analyses. Two powerful and commonly applied tools to perform such analyses are (isotope) hydrograph separation (IHS, Buttle, 1994; Klaus and McDonnell, 2013) and End Member Mixing Analysis (EMMA, Christophersen and Hooper, 1992; Hooper et al., 1990). These methods rely on the determination of 'end-members' which contribute to the mixture, in this case streamflow, and use the change in solute concentration(s) or isotopic composition(s), to calculate the contributions of the different end-members to streamflow. The chosen end-members typically include precipitation, groundwater and soil water, or 'pre-event' and 'event'-water, which represent the water previously stored in the catchment (i.e., groundwater and soil water) and the incoming precipitation, respectively. Both methods assume conservative, linear mixing between the end-members, and thus preclude the use of reactive tracers or time-variable end-member compositions.

However, the effective end-member composition might not be constant over time if we consider the spatial variability in the end-member composition in the catchment. Spatial variability in pre-event water can be large (Carey and Quinton, 2005; Kendall et al., 2001; Klaus et al., 2015b), and can result in temporal variability of the effective end-member composition if the relative contributions from different groundwater (and soil water) stores to streamflow change during the event (Burns et al., 2003; Inamdar and Mitchell, 2006; Jencso et al., 2010; McGlynn and McDonnell, 2003; Penna et al., 2016; Rinderer et al., 2019; von Freyberg et al., 2014). Usually, the pre-event water composition in small catchment studies is based on a pre-event streamflow (baseflow) sample, or on one or a few groundwater samples (Penna and van Meerveld, 2019). In such cases, these samples

are assumed to be representative for the pre-event water that is stored in the entire catchment (e.g., Brown et al., 1999; Hugenschmidt et al., 2014; Penna et al., 2015). These few samples likely lead to an underestimation of the total variability in the catchment. Baseflow might also not be a good representation of the catchment average groundwater if the relative contributions change during the event. Therefore, it is important to quantify the spatial variability, and to report the sensitivity and uncertainty related to source-area analyses.

There are two common ways to quantify the uncertainties related to source-area analyses. The first method is to use a Monte Carlo approach, in which the (pre-)event water composition is varied over the observed or estimated range (McDonnell et al., 1991; Rodhe, 1987). This method allows for the quantification of the 'sensitivity' of the analysis to an end-member characterization. The second method is through an estimation of the uncertainty of the mixing analysis with a gaussian error-propogation method, such as suggested by Genereux (1998). In this analysis the uncertainty depends on the differences in the absolute end-member concentrations (i.e., the difference between the pre-event and event water composition) and the variability within each end-member (i.e., the spatial or temporal variability of the pre-event water and event water). However, information on the spatial or temporal variability of end-members (such as groundwater) is often lacking, which likely results in an underestimation of the total uncertainty. In the case of groundwater this is especially problematic because the total uncertainty is most sensitive to the uncertainty of the component that contributes most to streamflow (i.e., the pre-event water (groundwater) in most undisturbed headwater catchments in temperate climates; Genereux, 1998).

More knowledge on the spatial variability in shallow groundwater composition allows us to learn more about the hydrologic functioning of different locations in the catchment, can aid a more refined representation of groundwater source-areas and flow pathways, and improve estimates of the uncertainties of source-area analyses. This will contribute to more robust conclusions about the relative contributions of groundwater (or different parts of the groundwater) to streamflow and catchment-scale

runoff generation processes. Adequate quantification and representation of the spatial and temporal variability in shallow groundwater chemistry in source-area analyses is paramount because the conclusions drawn from such analyses shape our conceptual understanding of catchments. This conceptual understanding is used in the development of modelling frameworks and inherently influences the way in which we study and interpret hydrological processes, even for large-scale catchments or future conditions.

This Ph.D. dissertation aims to obtain a refined representation of shallow groundwater contributions to streamflow at the headwater catchment-scale. More specifically, in this Ph.D. dissertation I address the knowledge gap regarding the spatial and temporal variability in shallow groundwater chemistry and the factors that cause it. I use spatially distributed hydrochemical observations for the identification of source-areas to streamflow and conceptualization of catchment-wide connectivity, and provide an estimate of the uncertainty in source-area analyses due to spatial variability in the composition of groundwater. In the thesis, I address the following research questions:

1. What is the spatial and temporal variability in shallow groundwater composition during baseflow conditions? And, can this variability be explained by topographic or hydrodynamic site attributes, or by subsurface characteristics?
2. Which (groundwater) sources contribute to baseflow, and which to stormflow?
3. What is the uncertainty in hydrograph separation analyses due to the spatial variability in shallow groundwater composition?

Chapter 2

Study Area

2.1 A short history

The work presented in this thesis was carried out in the Studibach, which is a headwater catchment located in the pre-alpine Alptal valley in Switzerland. Research in the Alptal started in 1963. Initially, measurements were conducted in eleven headwaters, but after a destructive storm the measurements were reduced to three subcatchments: Erlenbach, Vogelbach and Lümpenenbach (Figure 2.1). Hydrological research in the Alptal region was initially focused on the quantification of the influence of forest on hydrological processes, and later expanded to water quality, bedload transport, nitrate dynamics, runoff processes, and snow hydrology, with continuous measurements starting already in the late 70's (Hegg et al., 2006). Further development of the measurement and sampling schemes in the Erlenbach resulted in multi-year surveys of snow depth (Stähli and Gustafsson, 2006), streamwater chemistry (BAFU), and high-frequency measurements of streamwater and precipitation compositions at the Erlenbach outlet (Knapp et al., 2020; von Freyberg et al., 2017), amongst many other studies.

Research in the Studibach started with the installation of a hydrometric monitoring network, consisting of 51 groundwater wells and seven streamflow gauging stations in 2009/2010 (Rinderer et al., 2014). The

Studibach was selected for the study on groundwater because it is impacted less by ditching than the Erlenbach. However, Studibach neighbors Erlenbach so that the opportunity to combine long time series and high-frequency measurements (Erlenbach) with detailed spatial information (Studibach) still exists.

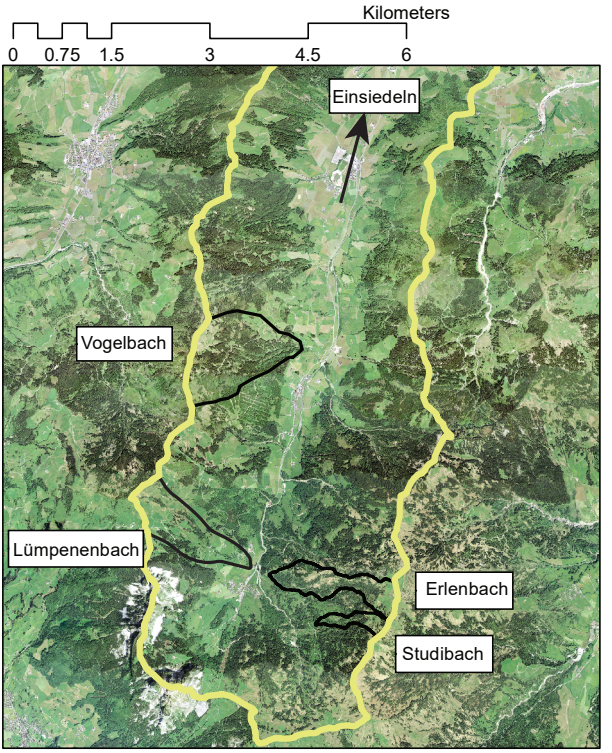

Figure 2.1 – Orthophoto of the Alptal valley and boundaries for the Vogelbach, Lümpenenbach, Erlenbach and Studibach (black lines) and the Alp (yellow line) catchments.

2.2 Studibach

The 20-ha Studibach is a pre-alpine headwater catchment in the Alptal (Figure 2.1). The Studibach is located roughly 40 km southeast of Zurich, Switzerland (N47.038°, E8.723°), and ranges from about 1270 m asl to 1650 m asl. The Alptal is typical for the Swiss pre-alpine region; it is characterized by frequent and high rainfall and low permeability soils, and therefore has mostly wet conditions. Mean annual precipitation amounts to 2300 mm y^{-1} and is relatively evenly distributed over the year. In the snow-free season (May-Oct) it rains approximately every other day. Stream and groundwater levels respond quickly to rainfall (within minutes to hours), and generally return to baseflow conditions within 1-2 days (Fischer et al., 2015; Rinderer et al., 2015).

The catchment is steep (average slope: 35°, Rinderer et al., 2014). The landscape is influenced by landslides and soil creep, resulting in a complex microtopography. Steep slopes of more than 20° are alternated with flatter areas. Many springs and stream heads can be found at the transition from concave to convex slopes, especially in the upper part of the catchment, and the natural drainage network is very dense (29 km km^{-2} van Meerveld et al., 2019). The steep terrain and high transport capacity of the streams have created channels with a step-pool morphology (Molnar et al., 2010) that cut into the alluvium and bedrock and have almost no riparian zone (Hagedorn et al., 2000). The steeper parts of the catchment and the ridge sites are covered by an open coniferous forest (*Picea abies* L. with an understory of *Vaccinium* sp., Hagedorn et al., 2000), whereas the flatter and concave parts of the catchment are moorlands or wet grasslands. The grassland areas in the upper part of the catchment are used for cattle grazing in the summer.

Soil depth ranges from 0.5 m at the ridge sites to more than 2.5 m in topographical hollows (Rinderer et al., 2014). At the grasslands and moorlands, where the water table is persistently close to the soil surface (Rinderer et al., 2014), mollic Gleysol, with a top soil that is high in clay

and carbonate and a permanently reduced Bg horizon occur (typically 43% clay, 42% silt and 15% sand Schleppi et al., 1998). The mostly forested ridge sites consist of an umbric Gleysol with an oxidized Bw horizon (49% clay, 46% silt, 5% sand) with macropores (Hagedorn et al., 2000; Schleppi et al., 1998). The underlying bedrock consists of three types of poorly permeable clay-rich Flysch: Ragazer Flysch, Schlieren Flysch and Wildflysch (Figure 2.2). Flysch is a reworked deep-water sedimentary rock, that is deposited in turbidites (submarine avalanches) on the continental shelf. Since these turbidites include weathered material from upslope, this results in highly heterogeneous deposits. The overall composition of these flysch formations are quite similar, although there are slight differences in the relative amounts of calcareous sandstone and argillite and bentonite schist layers (Mohn et al., 2000).

The Studibach was equipped with wells, stream gauges and precipitation gauges (Figure 2.2) in 2009/2010 to address questions regarding the spatial patterns of shallow groundwater dynamics (Rinderer et al., 2014, 2015), stream water chemistry (Fischer et al., 2015) and rainfall amount and isotopic composition (Fischer et al., 2016). Rinderer et al. (2014) showed that median groundwater levels are correlated to topographic indices, such as slope, curvature, Topographic Wetness Index (TWI) and upslope contributing area, and that the correlation differs depending on whether the indices are calculated for the site (local) or upslope contributing area. Rinderer et al. (2015) compared the groundwater response timing during rainfall events to rainfall intensity, antecedent moisture conditions, and the topographic site characteristics. They found that groundwater response timing was highly correlated to static topographic indices, and that rainfall intensity was more important than antecedent moisture conditions in determining the groundwater response timing. The median groundwater response time decreased with Topographic Wetness Index (TWI; Beven and Kirkby (1979)) for sites with TWI < 6, whereas topographically wetter sites responded almost immediately to rainfall. They also observed a decrease in the strength of the correlation between groundwater levels and TWI at the beginning of rainfall events, and an increase of this correlation after peak flow. For half of the events,

the groundwater levels at the wet sites close to the stream and in isolated depressions (44% of the catchment area) started to respond within 1 mm of rainfall, whereas a groundwater response occurred in 87% of the catchment after 5 mm of rainfall (Rinderer et al., 2015).

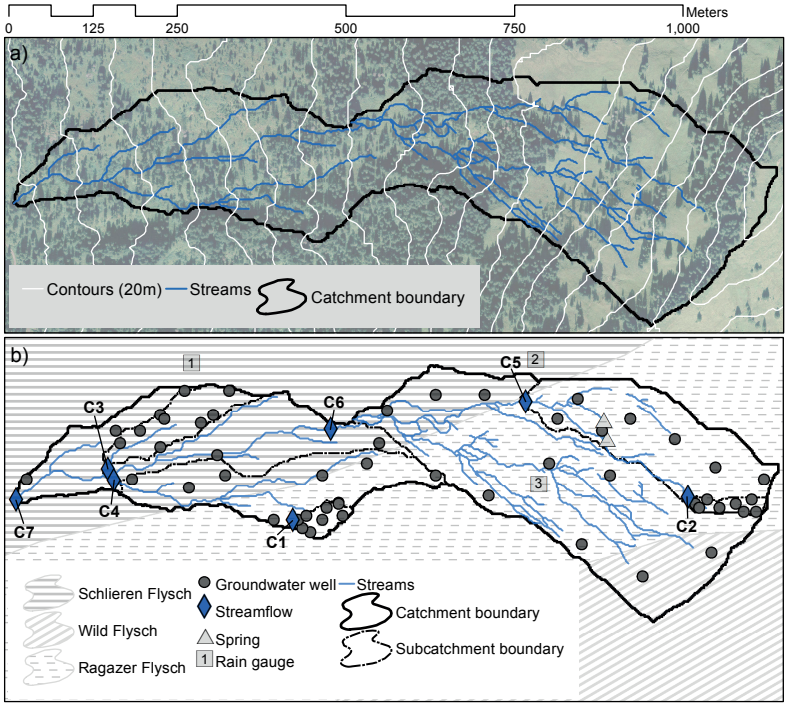

Figure 2.2 – Maps of the study area with a) an aerial photograph, contour lines (20 m interval), streams (blue), and catchment boundary (black), and b) the catchment (black) and sub-catchments (dark grey) boundaries, streams (blue) and streamflow gauges (C1-C7; blue diamonds), springs (light grey triangles), groundwater wells (dark grey dots), and rain gauges (light grey squares), as well as the three main types of Flysch (shading). Note that the stream network was manually surveyed for the upper part of the catchment but that this has not been completed yet for the lower catchment, giving a false impression of a lower drainage density in the lower catchment. Source geologic map (GK500) and aerial photograph (Aerial Images no. 20000090712703): Swiss Federal Office of Topography. Figure from Paper I.

Rinderer et al. (2019) further used the groundwater level data for the development and testing of a data-driven model to determine the fraction of the catchment area that is active or connected to the stream. This model uses the water level data from all 51 wells in the catchment and time series clustering to assign each pixel in the catchment to one of six groundwater clusters (based on TWI). For each time step, the average relative groundwater level for all monitored wells that belong to each cluster is calculated and assigned to all pixels with a similar TWI. This relative water level is then transformed into an absolute water level based on the correlation between soil depth and slope. If the water level is within 30 cm of the soil surface (the part of the soil where the hydraulic conductivity is high), the pixel is considered active, otherwise it is considered inactive. If a pixel is active and, based on surface topography, connected to the stream via other active pixels, it is considered to be connected to the stream. This results in groundwater activation maps (see example in Figure 2.3) through which the dynamic expansion and contraction of subsurface source areas can be studied. Rinderer et al. (2019) highlighted that a large part of the catchment is transiently connected to the stream, but also that there are isolated active zones that do not get connected to the stream during rainfall events.

Sampling of streamwater, springs and groundwater in the Zwäckentobel showed that baseflow consists mostly of flow from perennial groundwater springs, that are permanently connected to the streams (Fischer et al., 2015). The analysis of 13 rainfall-runoff events in six neighboring headwater catchments (of which the Studibach was one), showed that the pre-event water contributions to streamflow were largest for small events and that increasing precipitation resulted in a larger fraction of event water (Fischer et al., 2017). The latter agrees with the analysis of 24 rainfall events in the Erlenbach, from which von Freyberg et al. (2018) concluded that event water contributions were small and depended don the storm characteristics and catchment wetness.

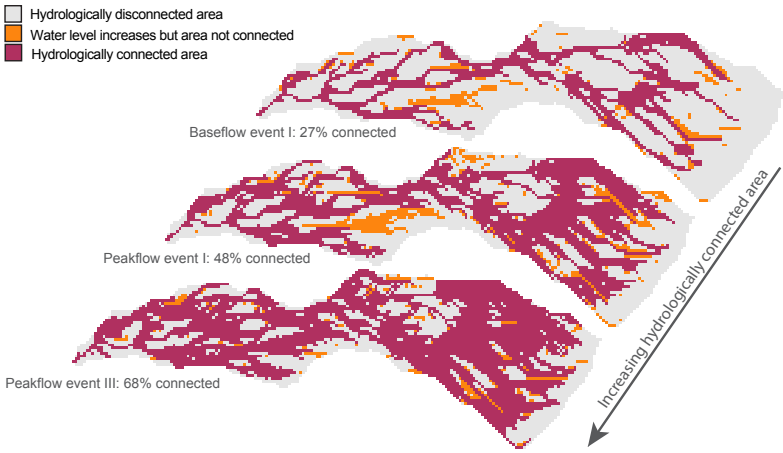

Figure 2.3 – The simulated spatial pattern of the hydrologically connected area for three different flow conditions: from relatively low flow (baseflow prior to event I; top), to intermediate flow conditions (peakflow during event I; middle), to the period of highest discharge for the studied events (peakflow during event III; bottom). See section 3.2 and Table 3.2 for information on the events. Grey indicates the hydrologically disconnected areas (water level more than 30 cm from the soil surface), red indicates the hydrologically connected area (i.e., water level within 30 cm from the soil surface and connected to the stream via other active areas), and orange indicates the active but disconnected area (i.e., water level in the upper 30 cm of the soil but not connected to the stream by other active areas). The connected area was simulated based on the measured groundwater levels and a data-driven model that uses surface topography to estimate the water level for unmonitored grid cells (cf. Rinderer et al., 2019). Figure from Paper II.

Chapter 3

Methods

In this chapter the methodology applied throughout this Ph.D. work is described. For detailed information that is not applicable to the core of the thesis, I refer to the corresponding journal publications.

3.1 Hydrometric observations

For all groundwater and streamwater level observations, I used the network of 51 groundwater wells and seven streamflow gauges that was installed in the Studibach in 2009/2010 (Figure 2.2). This hydrometric network distinguishes seven nested sub-catchments, ranging in size from ∼0.2 to ∼20-ha (Figure 2.2). Eight or nine groundwater wells were installed in each sub-catchment. The wells were hand-augered until the bedrock (0.5 to 2.5m), screened over the entire length and sealed with a bentonite clay layer. The wells were distributed based on their TWI and represent the range of dry and wet sites within each sub-catchment. They are numbered so that C1.1 is the well with the lowest TWI in the smallest sub-catchment, and C3.9 is the well with the highest TWI in one of the intermediately-sized sub-catchments. The groundwater levels were monitored in all the wells.

Stage height was measured at each of the sub-catchment outlets with an H-flume (C1 and C2), a weir (C3, C4 and C5) or directly in the river (C6

and C7). Stage height was converted to discharge with a stage height-discharge relation based on eight (the smallest catchments) to twenty (C6 and C7) salt dilution gaugings for each site.

All water levels were measured with water capacitance water level loggers (Odyssey Dataflow Systems Pty Limited) or pressure transducers (DCX-22 CTD Keller AG für Druckmesstechnick or STS DL/N 70, Sensor Technick Sirnach AG). The pressure readings were adjusted for changes in barometric pressure and temperature using the data from the MeteoSwiss station of Einsiedeln (910 m a.s.l; ca. 10 km distance from catchment outlet). Electrical conductivity (EC) was measured at all streamflow stations and ten to twenty groundwater wells. The EC measurements occurred over different periods for the different measurement sites, and did not follow a specific protocol. Precipitation was recorded at three locations in the catchment (Figure 2.2) with tipping bucket rain gauges (0.2-mm resolution, Odyssey Dataflow Systems Pty Limited).

3.2 Water sampling

I collected groundwater, soil water, rainwater and streamwater samples during different hydrological conditions to study the composition of these water sources, and to determine their relative fractions in the streamflow mixture.

I sampled during baseflow conditions to quantify the spatial variability within different water compartments, to identify groundwater areas based on their composition, and to estimate the contributions from different groundwater areas to streamflow. The sampling at baseflow conditions took place during nine snapshot sampling campaigns in the snow-free season between June 2016 and November 2017 (Table 3.1). During each snapshot campaign, I collected groundwater samples from all wells with a sufficient groundwater volume (34 to 47 wells for the different campaigns) and streamwater samples at the outlets of the seven sub-catchments. All wells were purged the day before sampling by

pumping them dry or extracting at least two times the well volume. For all campaigns except the first, I also collected soil water samples from twelve to eighteen suction lysimeters at up to six sites (three forested, three non-forested), at three different elevations in the catchment (1361, 1502, 1611 m a.s.l.) at 15, 30 and 50 cm below the soil surface. The suction lysimeters were set to a tension of 50 mbar the day before sampling to extract the soil water.

Table 3.1 – Overview of the conditions prior to the nine snapshot sampling campaigns (SC): the streamflow at the outlet of C4 and C5 at 8:00 am on the day of sampling (Q_{C4} and Q_{C5}), the sum of precipitation in the 7 days (P_7) and 2 days (P_2) prior to the sampling campaign, the average temperature in the week before the sampling campaign (T_7), and the number of wells that contained water and could thus be sampled. The streamflow of sub-catchment C4 and C5 was used as a proxy for catchment wetness because these stream level data series are most complete and the rating curves at these V-notch weirs are most reliable.

SC	Date	Q_{C4} $Ls^{-1}km^{-2}$	Q_{C5} $Ls^{-1}km^{-2}$	P_2 mm	P_7 mm	T_7 °C	n_{wells}
1	20 Jun 2016	38	60	134	27	11.3	42
2	19 Jul 2016	28	4	96	0	13.6	34
3	31 Aug 2016	5	< 3	23	23	19.1	38
4	05 Oct 2016	3	4	11	0.6	12.2	37
5	31 Oct 2016	3	5	51	0.1	6.6	39
6	17 May 2017	95	24	37	0.3	12.8	36
7	20 Jun 2017	4	4	37	0	18.4	38
8	24 Aug 2017	–	5	16	1.0	16.0	35
9	12 Oct 2017	–	< 3	29	0.1	8.3	35

In addition, I collected streamwater, rainwater and groundwater samples during rainfall events to investigate the sources contributing to streamflow during events, as well as the changes in groundwater compositions over time. I collected samples of precipitation at up to three locations, streamwater at up to three locations, and groundwater at up to five locations, for a total of eight events. I used the streamwater and rainwater samples of five events for the subsequent analyses (I-IV in Table 3.2) in Paper II and Paper III. The other events were not used because the event was not captured properly (i.e., missing either the beginning or the end of the event), because snowfall occurred during the event, or because there was no baseflow snapshot campaign available to compare the event sampling with. Potential further use of the groundwater samples taken during rainfall events is discussed in chapter 9.

Table 3.2 – Overview of the four events presented in this thesis: event duration (D, h), rainfall amount (P, mm), average and maximum 10-min rainfall intensity (I_p and I_{p-max}, mm h^{-1}), the range in specific discharge (ΔQ, mm h^{-1}, the maximum change in isotopic composition of the stream water (δ^2H, ‰), and the minimum and maximum simulated fraction of the catchment that was connected (A_{min}-A_{max}; cf. Rinderer et al. (2019)) during the event, and the date of the corresponding groundwater and soil water sampling campaign (Date SC).

Event	I	II	III	IV
Start date	02 Oct 16	25 Oct 16	03 Oct 17	05 Oct 17
D [h]	14	28	7	27
P [mm]	17	33	27	32
I_p [mm h^{-1}]	1.2	1.2	3.9	1.2
I_{p-max} [mm h^{-1}]	7	13	24	10
ΔQ [mm h^{-1}]	0.02 – 0.07	0.02 – 0.17	0.08 – 0.43	0.07 – 0.30
$Q_{min}[\delta^2H$ ‰]	-70.5	-75.3	-73.7	-69.1
$Q_{max}[\delta^2H$ ‰]	-65.7	-67.6	-69.1	-65.2
A_{min}-A_{max} [-]	0.27 – 0.48	0.27 – 0.35	0.33 – 0.68	0.33 – 0.67
Date SC	05 Oct 16	05 Oct 16	12 Oct 17	12 Oct 17

For the groundwater and streamflow sampling during rainfall events, I used automatic samplers (full-size portable sampler, 6712, ISCO Teledyne, USA). The sampling interval was adjusted according to the predicted

event duration. The samplers were set to a timer if the predicted time of onset of the event was during the night. I applied a multi-interval program to sample streamwater every ten to twenty minutes at the start of the rising limb (restricted to six samples). The remaining eighteen samples were taken at an hourly-interval. For the precipitation sampling, I used passive sequential rainfall samplers that collect a sample approximately every 5 mm of rainfall (built after Kennedy et al. (1979); detailed description in Fischer et al. (2019)). All samplers were emptied within 24 hours after sample collection to avoid any fractionation.

All analysed events were larger than the average daily precipitation (median: 10.0 mm; mean ± sd: 14.1 ± 13.8 mm) and within the upper 30% of daily precipitation based on data from the long-term meteorological station Erlenhöhe, located 500 meters from the catchment outlet (including all rainy days with more than 1 mm precipitation; 7452 days in 36 years, 1981-2017 period; Stähli, 2018). Total precipitation and duration of the four events ranged from 17 to 33 mm and from 7 to 27 hours (Table 3.2). The average and maximum 10-minute rainfall intensity ranged between 1.2 and 3.9 mm h^{-1} and between 4.8 and 22.8 mm h^{-1}, respectively. Discharge at the catchment outlet increased least (0.02 to 0.07 mm h^{-1}) for the smallest event (I), and most for event III (0.08 to 0.43 mm h^{-1}).

The simulated fraction of the hydrologically connected area for the events varied from 0.27 (before the start of event I) to 0.68 (during peak flow of event III) of the total catchment area (methodology cf. Rinderer et al. (2019); see Figure 2.3). For event I, the hydrologically connected area extended laterally from the stream up, but remained confined to the flat ares. For event II the expansion of the hydrologically connected area was minimal. For the intermediate events, the lateral extension was larger and parts of the hillslopes became connected. However, simulations of the data-based model suggested that during all events, large parts of the catchment remained hydrologically disconnected.

3.3 Sample treatment and analysis

The Electrical conductivity (EC) of all water samples was measured directly in the field using a Multi 3420 conductivity probe (WTW Measurement Systems Inc). The samples were collected in polyethylene bottles (50-300 mL) and stored at 4°C until preparation and analysis within a week after collection, except for SC8 (August 2017) for which the samples were frozen (-18 °C) after collection until sample preparation a few days before analysis. Sample preparation involved filtering (0.45 μm, SimplepureTM Syringe Filter) and redistribution of the sampled water into three different vials to measure cation and anion concentrations and the isotopic composition. The samples for cation analysis were acidified with 50 μL of 50 mM $HNO_3{}^-$ solution to mobilize trace metals and reduce microbial activity.

The cation and anion analyses were performed at the Physics of Environmental Systems laboratory at ETH Zurich, using an ion chromatograph (861 Advanced Compact IC, Metrohm) for anions and a mass-spectrometer (ICP-MS 9700, Agilent technologies) for cations, except for the first campaign (SC1), for which the samples were analysed on an emission-spectrometer (ICP-OES 5100, Agilent technologies). Calibration curves were obtained from measurements immediately before or after the sample analyses with five calibration standards. Stable water isotope analyses were performed with a Cavity Ring-Down Spectrometer (L2140-i or L2130-i Picarro Inc.) at the isotope laboratory of the Chairs of Hydrology at the University of Freiburg with a reported precision of $\pm$0.16 ‰ for δ^{18}O and $\pm$0.6 ‰ for δ^{2}H. I use the standard delta notation (‰) for isotopic compositions and excess-Deuterium (D_{ex}) throughout the thesis.

3.4 Drivers of spatial variability

To improve our understanding of the observed spatial variability in shallow groundwater composition, I investigated the correlation between solute concentrations and site attributes describing the local topographic and hydrodynamic conditions, subsurface topography, and the leaching potential of different soil and bedrock samples. I used topographic and hydrodynamic site attributes to describe the potential for water to be accumulated or drained from a specific location and did electrical resistivity (ERT) profiling to find the depth-to-bedrock at different locations. I also investigated the mobile elements for five soil and bedrock samples by performing leaching experiments.

3.4.1 Topographic and hydrodynamic site attributes

The topographic and hydrodynamic site attributes were calculated for each location separately. The topographic attributes were based on a 6 by 6 meter Digital Terrain Model (DTM) derived from LiDAR data. The topographic attributes included: slope (SL), TWI, plan curvature (PLC) and profile curvature (PRC), downslope index (DI), ruggedness (RUG), and gradient to creek (GTC), as well as attributes that describe the upslope flow pathways, such as average slope of the upslope flowpath (ASL) and maximum upslope flowpath length (AFL). All topographic attributes were computed using the Whitebox Geospatial Analysis Tools (GAT Lindsay, 2016), using a channelization threshold of 25 to identify the stream cells. The hydrodynamic attributes were based on the groundwater levels measured between 1 December 2015 and 29 November 2017, aggregated to hourly values to reduce data noise. The hydrodynamic attributes included skewness of the frequency distribution of the groundwater levels (Skew), median groundwater level (Median) and fraction of time that the groundwater level was within 10 or 30 cm of the land surface (PST10 and PST30). See Paper I for the complete list of all hydrodynamic and topographic attributes.

First, the explanatory power of hydrodynamic and topographic site characteristics for the variation in groundwater composition was explored using Spearman's Rank correlation analysis. Then, a Random Forest (RF) analysis was applied to quantify the relative importance of the site attributes. RF is a classification algorithm that uses random subsets of the data to construct multiple decision trees. Finally, a Principal Component Analysis (PCA) was used to compare site attributes that were most important in the RF to the solute concentrations. PCA is a statistical method in which a new projection (i.e., a new coordinate system) is defined for the data in such a way that the first component describes most of the variability, and any following components are orthogonal to its preceding component. The isotopic composition was not considered in this analysis due to seasonal changes in the composition of precipitation, and thus groundwater. All calculations were performed on log-transformed data, and for the PCA the data were additionally z-transformed.

3.4.2 Subsurface topography

To study the importance of subsurface topographical features for the spatial distribution of groundwater composition, I measured subsurface resistivities along nine profiles using Electrical Resistivity Tomography (ERT). These measurements were performed in collaboration with Adrian Wicki (Swiss Federal Research Institute WSL). We conducted measurements on 14 June, 21 August and 22 August 2018, and focused on both forested and non-forested areas. We selected key locations in the catchment that are representative for the dominant land-forms, such as the upper spring zone, and several ridge features (Figure 3.1). The profiles were performed parallel (n=6) or perpendicular (n=3) to the slopes, except for profile 6, where the profiling was performed in both directions. The ERT measurements were taken with a Syscal Pro (IRIS Instruments) ERT profiling system, using 48 elektrodes in Wenner-Schlumberger arrays at 0.8 m spacing, resulting in profiles of 37.6 m length. The profiles were inverted using RES2DINV.

3.4.3 Soil and bedrock leachates

To investigate how the composition of soils and bedrock are related to the observed groundwater compositions, I performed leaching experiments on three soil samples and two bedrock samples in collaboration with Dr. Christophe Hissler at the Luxembourg Institute of Science and Technology (LIST). The experiments were performed on soil samples from 30 cm depth in a riparian zone and 30 cm and 5 cm depth on a hillslope, and on bedrock samples from a rather thickly banked sandy carbonate bedrock and a thinly banked, silty carbonate bedrock (Figure 3.1). All laboratory methods were conform to Hissler et al. (2015). The measurements were performed on 63 μm-sieved material from which one gram of material was isolated. The sequential leaching was performed in two stages of increasing acidity, using first 0.05 N HAc and afterwards 1 N HCl as a solvent. Major and trace element concentrations were determined using Inductively Coupled Plasma Mass Spectrometry (Perkin-Elmer DRC-e).

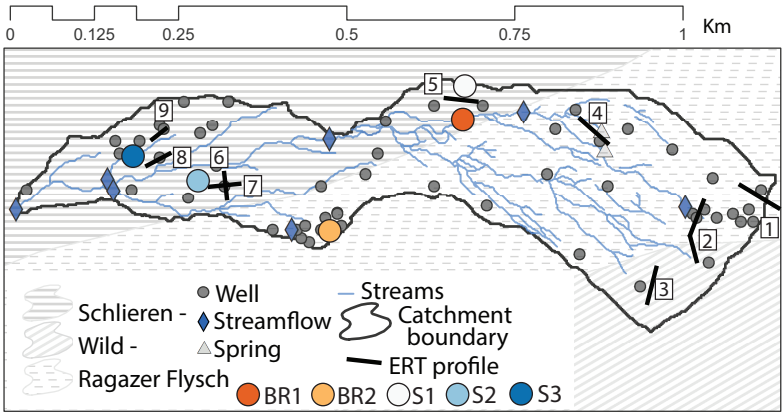

Figure 3.1 – Map of the Studibach catchment with the location and number of each ERT profile (black lines, 1-9), the sampling locations of soils and bedrock for the leaching experiments (colored circles: thickly banked bedrock (BR1, dark orange), silty bedrock (BR2, light orange), forest soil (S1, white), hillslope topsoil (S2, light blue), and riparian soil (S3, dark blue), and wells (grey dots), streamflow gauging sites (blue diamonds) and springs (grey triangles). The shading represents the three different types of flysch bedrock.

3.5 Data evaluation

The normality of the frequency distributions of the concentrations, EC, δ^2H and δ^{18}O was tested using the Shapiro-Wilk test. Only EC, δ^2H, δ^{18}O, and the concentrations of calcium and lithium were approximately normally distributed, the data for all other elements were log-transformed prior to the statistical analyses. A p-value of 0.05 was used for all statistical analyses. All data analyses were performed with R-studio (R Core Team, 2013).

3.5.1 Delineation of landscape elements

The temporal variability for each parameter (i.e., solute concentrations, EC, and isotopic composition) was determined by calculating the average concentration across the catchment for each campaign (μ), then, selecting the maximum average concentration and minimum average concentration, and lastly, dividing this range in average concentration by the average of all groundwater concentrations measured during all snapshot campaigns.

The persistence of the spatial variability of the concentrations was then assessed by calculating the mean relative difference (M_{RD}) for each parameter and each well. This was done using equation 3.1, where x_j is the measured value for that well during sampling campaign j, μ_j the average value for all wells sampled during sampling campaign j, σ the standard deviation for all wells during sampling campaign j, and n_{sc} is the number of sampling campaigns for which there were samples for that well.

$$M_{RD} = \frac{1}{n_{sc}} \sum_{j=1}^{n_{sc}} \frac{x_j - \mu_j}{\sigma_j} \tag{3.1}$$

The average spatial variability (M_{RD}) was used to delineate different landscape elements through k-means clustering. The Calinski-Harabasz (CH) criterion indicated three to five clusters would be optimal; four clusters were selected because of the relatively even distribution of wells per cluster (14, 11, 9 and 7 wells, respectively).

3.5.2 Hydrograph separation

Isotope hydrograph separation was used to calculate the fractions of event water (f_e) and pre-event water (f_{pe}) in each stormflow sample. This calculation is subject to five underlying assumptions:

1. The isotopic content of the event and the pre-event water are significantly different.
2. The event water has a constant isotopic signature in space and time, or any variations can be acounted for.
3. The pre-event water has a constant isotopic signature in space and time, or any variations can be acounted for.
4. Contributions from the vadose zone are negligible, or the isotopic signature of the soil water must be similar to that of groundwater.
5. Surface storage contributes minimally to the streamflow.

The pre-event water fraction (f_{pe}) was calculated using equation 3.2:

$$f_{pe} = \frac{C_t - C_e}{C_{pe} - C_e} \tag{3.2}$$

in which f_{pe} is the pre-event water fraction, and C_t, C_e and C_{pe} are the compositions of streamwater, event water and pre-event water, respectively. The incremental weighted mean of rainfall was used to characterize the event-water composition (C_e), C_{pe} was based either on a baseflow sample, or the average of one or multiple groundwater samples (see section 3.5.5). All calculations were performed using δ^2H as the tracer.

3.5.3 End-Member Mixing Analysis

End Member Mixing Analyses (EMMA; Christophersen and Hooper, 1992; Hooper et al., 1990) were used to investigate the contributions of rainfall, soil water and different types of groundwater to streamflow. The first two principal components were used to calculate the fraction of each end-member in streamflow. For the PCA during baseflow conditions, the measurements of the isotopic compositions (δ^{18}O, δ^{2}H and D_{ex}), and solute concentrations (Ca, Mg, Na, SO_4, Cl, K, Zn, Cu, Fe and Mn) of groundwater and streamwater sampled during each campaign were used as input data. For the EMMA during events only conservative tracers were used. A tracer was considered conservative if it had a linear correlation with at least one other tracer. The groundwater, soil water and streamwater samples were used to determine the conservativeness, and the threshold for a linear correlation was set to $R^2 \geq 0.5$ and $p \leq 0.05$ (i.e., following the threshold of Barthold et al. (2011)). Using this threshold, δ^{18}O, δ^{2}H, Ca, EC, Mg and Ba were considered conservative. The uncertainties on the mixing fractions were calculated using the method of Genereux (1998) (detailed methodology in Paper III) .

3.5.4 Deviation from conservative mixing

To test if it is possible to identify changes in hydrological connectivity from streamwater solute concentrations, I investigated when solute concentrations deviate from a concentration that is expected based on conservative mixing of rainfall and baseflow. To this end, I compared the measured stormflow concentrations to the solute concentrations that I estimated using the pre-event water fraction (derived from two-component hydrograph separation), and the concentrations for this solute in baseflow and rainfall, following equation 3.3:

$$C_{es_x} = (C_{BF_x} \cdot f_{pe}) + (C_{P_x} \cdot (1 - f_{pe})) \tag{3.3}$$

in which C_{es_x} is the estimated concentration for solute x, C_{BF_x} and C_{P_x} are the concentrations for solute x in baseflow and rainfall, and f_{pe}

is the pre-event water fraction for that sample, as calculated with the two-component hydrograph separation based on δ^2H (Equation 3.2).

The difference between the measured C_{Q_x} and estimated concentrations $C_{Q_e s_x}$) was compared to specific discharge measured at the catchment outlet and to the characteristic concentrations in different source areas. I assumed that overestimations or underestimations of the concentrations ($C_{es_x}/C_{Q_x} < 1$ or > 1, respectively) indicate contributions from source areas that were not active during baseflow, and that have a lower or higher concentration than the sources that contributed to baseflow respectively, or that these deviations are the result of reactive transport processes (see Section 5.1.

3.5.5 Uncertainty estimations hydrograph separation

To quantify the sensitivity of the hydrograph separation to different pre-event water characterizations, I compared the event-average pre-event water fractions using five ways to characterize C_{pe}:

1. a pre-event streamflow sample, or a baseflow sample taken during the snapshot campaign
2. the composition of a (riparian) groundwater well (n=11 for riparian wells, and n=38 for all groundwater wells)
3. the average composition of all riparian groundwater wells
4. the composition of all groundwater wells
5. one, three, six or nine randomly selected (riparian) groundwater wells

For the second characterization, the hydrograph separation was repeated 11 times if the focus was on only the riparian groundwater wells and 38 times if the focus was on all groundwater wells. For the last characterization the hydrograph separation was repeated 1000 times for different randomly selected wells (see detailed methodology in Paper III). Any computation for all groundwater wells also includes all riparian

groundwater wells.

To estimate the uncertainty in the calculated pre-event water fractions (W_{fpe}), a Gaussian error-propagation method was used (Equation 3.4, after, Genereux, 1998). The standard deviation in rainfall (sampled during the event) was used as the uncertainty of the event-water composition (W_e), and the laboratory accuracy as uncertainty of the streamflow composition (W_t). For pre-event water characterizations based on one sample, the parameter uncertainty was set to the laboratory accuracy, and when multiple wells were (randomly) selected to characterize the pre-event water, I used the standard deviation of these wells as W_{Cpe}.

$$W_{fpe} = ((\frac{C_e - C_t}{(C_e - C_{pe})^2} W_{Cpe})^2 + (\frac{C_t - C_{pe}}{(C_e - C_{pe})^2} W_{Ce})^2 + (\frac{-1}{(C_e - C_{pe})^2} W_{Ct})^2)^{\frac{1}{2}} \tag{3.4}$$

Finally, I estimated how much the calculated average (riparian) groundwater composition is affected by the number of sampled wells that are used in the analysis, and how many samples are needed to obtain an estimate of the average composition and variability of the groundwater composition. I did this by randomly selecting a number of wells (without replacement) and calculating the average and standard deviation of δ^2H and EC for the selected samples. The selections were based on a constrained random approach (i.e., the values were randomly selected measurements from the wells, but the locations of the wells was based on their topographic characteristics). This was repeated for all possible numbers of wells (1 to 38), and for all possible number of riparian wells (1 to 11). The standard deviation of the average pre-event water composition for 1000 realizations (i.e., 1000 random selections of wells for each set of number of wells) differed less than 0.01 ‰. Therefore the analysis was limited to 1000 realizations, rather than computing all possible combinations (e.g., 3.5 x 10^{10} combinations are possible when sampling 19 out of 38 wells).

Chapter 4

Spatial variability in shallow groundwater chemistry and its drivers

4.1 Spatial variability in groundwater chemistry

The chemistry of the shallow groundwater was dominated by calcium; its relative contribution to the electrical balance was larger than 80% for 94% of the samples (median 93%). The other most abundant solutes were bicarbonate, sulfate, magnesium, sodium, chloride, potassium, strontium and manganese (median concentrations $> 100\ \mu g L^{-1}$, listed in order of decreasing importance). Generally, solute concentrations increased during the summer when water levels and overall catchment wetness decreased, and concentrations were lowest directly after the snowmelt period. The isotopic composition of the groundwater during the sampling campaigns suggested three significantly different periods: shortly after snowmelt, characterized by isotopically more depleted water (SC1 and SC6), late summer, dominated by isotopically more enriched water (SC3, SC4 and SC8) and a mixture of both (SC2, SC5, SC7 and SC9).

The spatial variability in shallow groundwater chemistry was large, with concentration ranges spanning up to five orders of magnitude for elements such as calcium, manganese and zinc. For almost all parameters the median spatial variability (i.e., the median difference between the

minimum and maximum concentrations measured during a single sampling campaign; n=9) was larger than the temporal variability (i.e., the median difference between the minimum and maximum concentrations measured for a single well; n=47) (Figure 4.2). Exceptions were phosphate and aluminum, for which the temporal variability was larger than the spatial variability, and $\delta^{18}O$ and $\delta^{2}H$, for which the spatial and temporal variability were similar.

4.2 Delineation of landscape elements

Clustering the wells based on the average differences in concentrations from the catchment mean (i.e., M_{RD}; equation 3.1) yielded four shallow groundwater types (Figure 4.3).

- Type I: these wells are characterized by high concentrations of iron, manganese, and cobalt (transition metals). They are predominantly located in flat, wet areas, and have groundwater levels which are persistently close to the soil surface. These sites are further referred to as 'riparian-like';
- Type II: these wells are characterized by high concentrations of copper, lead, zinc and potassium (heavy metals), and are mainly located on steep slopes and ridges in the upper parts of the different subcatchments;
- Type III: these wells have above average concentrations of magnesium and sulfate, low transition metal concentrations, and a relatively depleted isotopic signature. These wells are relatively deep (110 ± 30 cm) and located mainly in the lower part of the catchment;
- Type IV: these wells have high concentrations of sulfate and magnesium, and low concentrations of transition metals and heavy metals. The wells are located in three specific locations close to bedrock outcrops with a contrasting lithology.

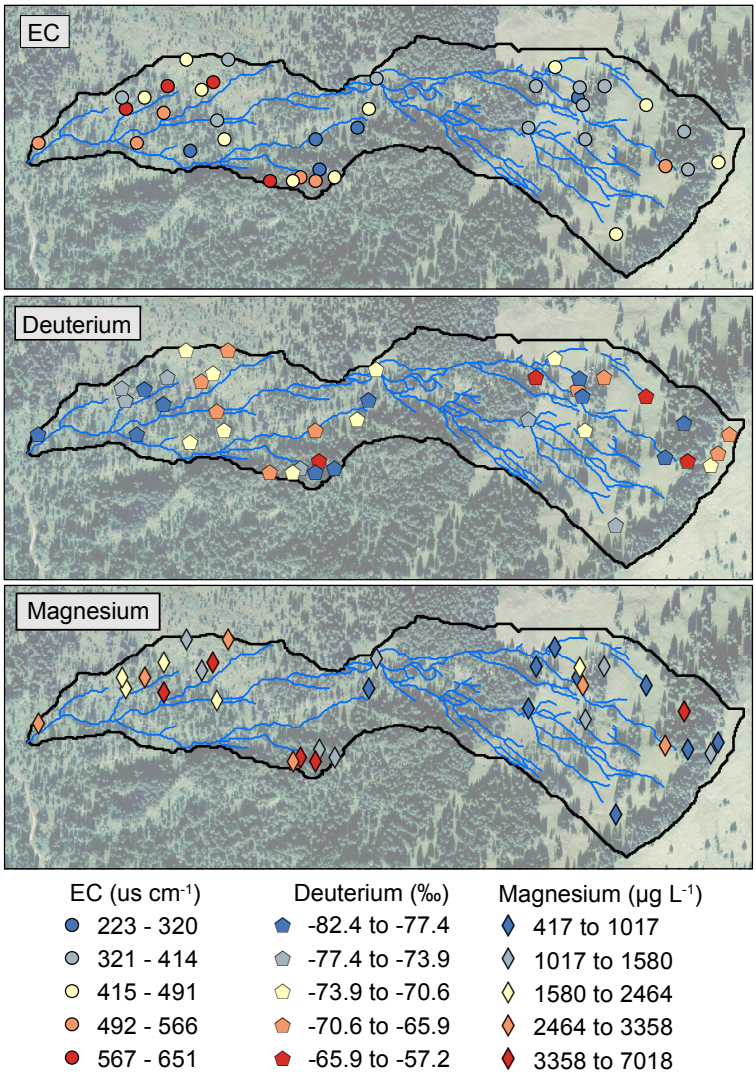

Figure 4.1 – Maps showing the spatial distribution in the EC (top, μS cm^{-1}), δ^2H (middle, ‰) and magnesium concentrations (bottom, μg L^{-1}) in shallow groundwater for the snapshot campaign on 5 October 2016 (SC4).

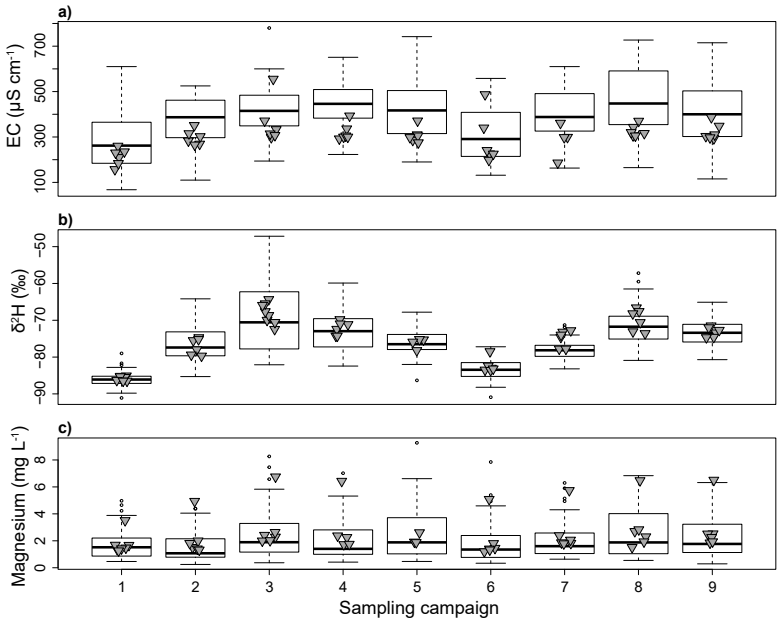

Figure 4.2 – Boxplots of a) EC (μS cm^{-1}), b) δ^2H (‰), and c) magnesium concentrations (mg L^{-1}) in shallow groundwater and the concentrations in streamwater (grey reversed triangles) for all nine sampling campaigns (SC1 to SC9). The box represents the 25th-75th percentile, the thick line the 50th percentile, the whiskers extend to the 25thpercentile – 1.5*interquartile range and 75th-percentile + 1.5*interquartile range, and the dots represent the outliers. Figure from Paper I.

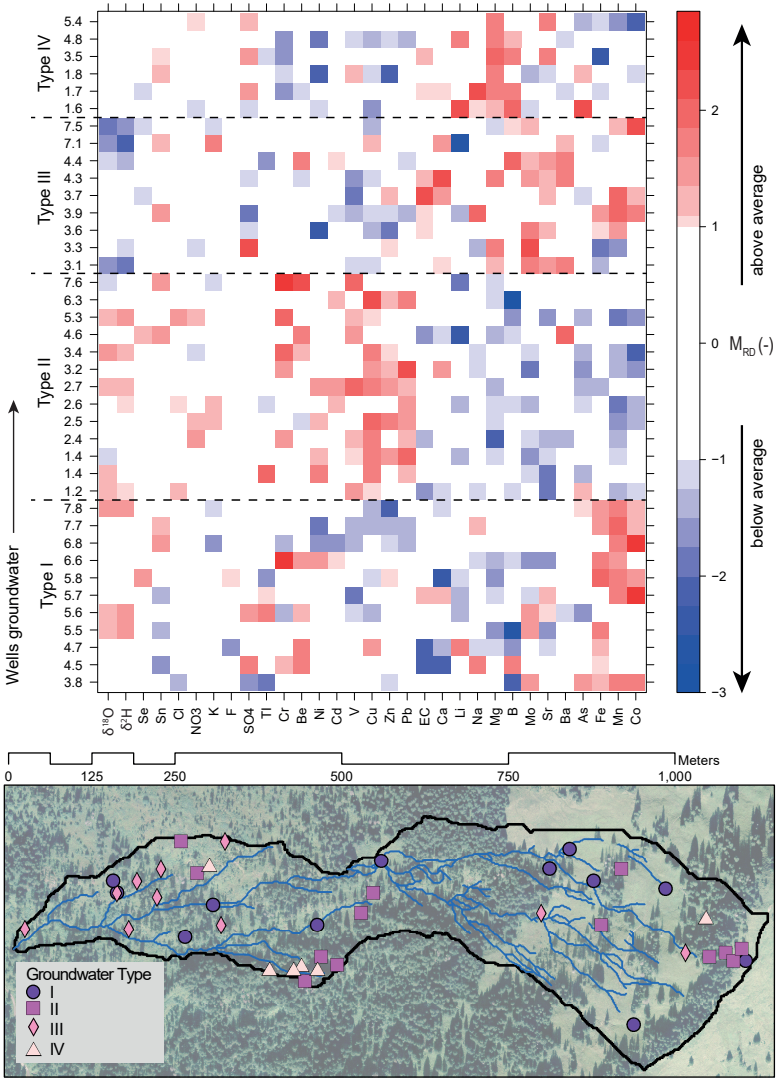

Figure 4.3 – (top) Mean relative differences (M_{RD}; Equation 3.1) from the catchment average concentrations in groundwater for each well (rows with well number) and parameter (columns). Wells are ordered according to the four groundwater types based on the M_{RD} values; (bottom) map with wells color coded according to the groundwater types. Figure from paper I.

4.3 Topographic and hydrodynamic site attributes

The extent to which the variability in solute concentrations could be explained by the hydrodynamic and topographic site attributes depended on the element of interest. Spearman's rank correlations ranged from ∼0 to 0.67 when all campaigns were analysed together, and were generally higher for the individual sampling campaigns (range: ∼0 to 0.81). Concentrations of heavy metals (copper, zinc, nickel, lead) were higher at sites that were relatively dry, whereas transition metal concentrations (iron, manganese, cobalt) were highest at sites that were predominantly wet. Potassium concentrations were strongly correlated with the site attributes and were highest at relatively dry (forested) sites. Correlations between site attributes and any of the other abundant solutes or EC were weak.

The average values of some site attributes varied significantly between some groundwater types, particularly these of Type I and II (riparian zone and hillslopes, respectively; ten of the eleven significantly different median attribute values; Table 4.1). For instance, the groundwater levels at Type I wells were negatively skewed and close to the surface whereas groundwater levels at Type II wells were positively skewed and further from the soil surface. Site attributes for Type III and IV were usually not significantly different from the other groundwater types, except for Skew, DI and SL, for which Type III was significantly different from Type II, and TWI, for which Type I and III were significantly different.

Table 4.1 – Median values for the hydrodynamic (top) and topographic (bottom) attributes that were significantly different between the four groundwater types, and the corresponding Kruskal-Wallis p.value (Median: median groundwater level; Skew: skewness of the groundwater level; PST30 and PST10: persistency of the groundwater level within the upper 30 or 10 centimeter of the soil profile; DI: downslope index; TWI: topographic wetness index; GTC: gradient to creek; PLC: plan curvature; PRC: profile curvature; SL: slope; RUG: ruggedness). Significantly different values for groundwater types are indicated with different letters ($^{a-b}$). See paper I for a detailed description of the site attributes.

Type	Hydrodynamic attributes			
	Median	Skew	PST30	PST10
I	-9.7^{a}	-1.9^{a}	0.85^{a}	0.51^{a}
II	-50.8^{b}	0.4^{b}	0.06^{b}	0^{b}
III	-27.4^{ab}	-0.6^{a}	0.59^{ab}	0.05^{ab}
IV	-24.7^{ab}	-0.4^{ab}	0.73^{ab}	0.01^{ab}
p.value	0.006	0.005	0.012	0.003

Type	Topographic attributes						
	DI	TWI	GTC	PLC	PRC	SL	RUG
I	29.0^{a}	6.6^{a}	9.6^{a}	4.9^{a}	0.3^{a}	14.8^{a}	1.4^{a}
II	14.5^{b}	4.4^{ab}	26.1^{b}	-0.9^{b}	-0.8^{b}	25.5^{b}	2.7^{b}
III	23.0^{a}	4.3^{b}	15.1^{ab}	2.9^{ab}	0.9^{ab}	15.6^{a}	1.7^{a}
IV	19.6^{ab}	4.6^{ab}	20.2^{ab}	2.4^{ab}	0.5^{ab}	21.0^{ab}	2.0^{ab}
p.value	<0.001	0.049	0.015	0.003	0.019	0.001	0.001

4.4 Subsurface topography

The soil and bedrock resistivities ranged from 20 to 3000 Ω m (Table 4.2). The resistivities were lowest at locations with wet or saturated soils and at the interpreted soil/bedrock interface. At the soil/bedrock interface the low resistivies likely reflect subsurface ponding (i.e., the perched groundwater). The highest resistivities were generally measured in sections along the profile that were notably dry, in the vicinity of trees or covered with tall grasses, or that were below the interpreted soil/bedrock interface.

The estimated depth to bedrock ranged from 0.5 to 3.5 m for all profiles (Table 4.2). On average, the estimated depth to bedrock was smallest in profile 1 (at the upper catchment boundary) and profile 5 (a profile in the forest), although for each profile there was a section with shallow soil (0.5-1 m). The interpreted bedrock surface was smoothest in the profile at the upper spring region (profile 4, Figure 4.4), in which the soil depth gradually increased downslope from the spring zone ($\sim$0 m) to a maximum depth of $\sim$3.5 m, and at profile 5, which was one of the two profiles in the forest. The bedrock topography at the other profiles was rugged and not always parallel to the bedrock topography. For instance, the estimated depth to bedrock in profile 2 (Figure 4.4) was largest (i.e., deepest soils) below the ridge, and shallowest below the topographic hollows. Similarly-sized undulations ($\sim$1 m) were present in all profiles, except profiles 4 and 5. Soils were marginally deeper in the lower part of the catchment (maximum depth for each profile: 3-3.5 m), than in the upper part of the catchment, where the maximum depth for three out of five profiles was less than 2.5 m. Steep profile sections with shallow to intermediate depths to bedrock co-occurred most frequently with Type II wells (hillslopes). Type I and Type III wells were located at locations with shallow bedrock depths where the slope changed from convex to concave, at locations with deeper soils, and in topographic hollows.

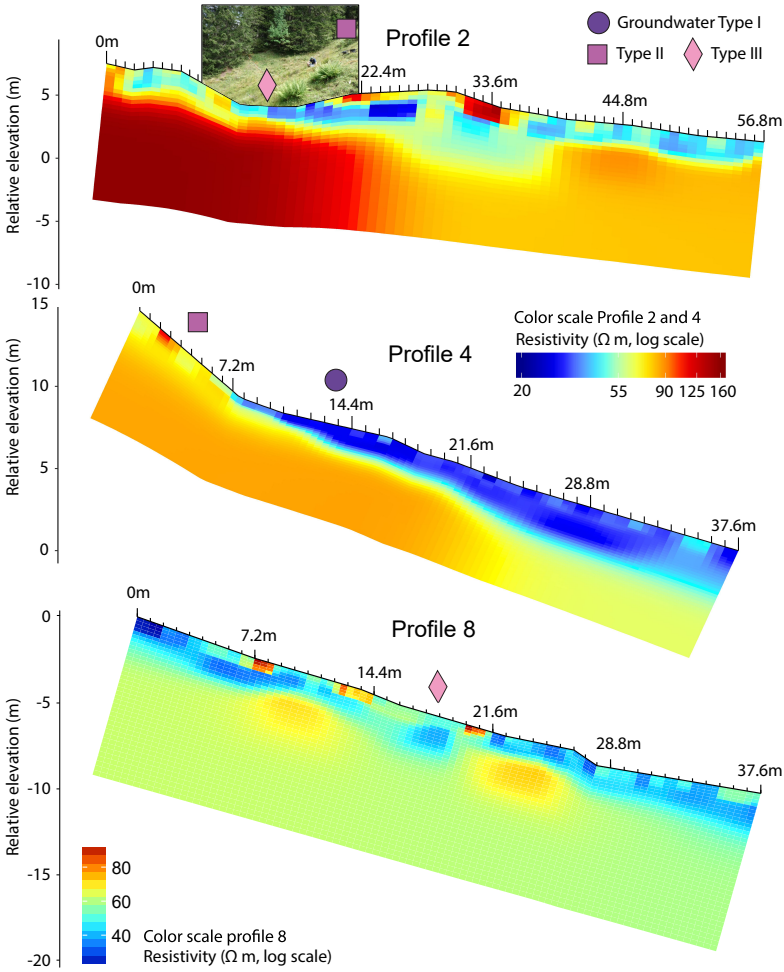

Figure 4.4 – The Electrical Resistivity Tomography (ERT) profiles measured at the upper part of the catchment (top, profile 2), the upper spring zone (middle, profile 4) and the lower part of the catchment (bottom, profile 8). Red indicates high resistivities and blue low resistivities (Ω m). Please note that the color scale used in profile 2 and 4 is different than the color scale used in profile 8. Groundwater types of nearby wells are indicated with a dark purple dot (Type I), purple square (Type II) or pink diamond (Type III).

Table 4.2 – Summary of the profiles: minimum and maximum depth to bedrock (D_{min}-D_{max}) and resistivities (R_{min}-R_{max}), forest cover, and orientation with respect to the overall catchment slope (parallel or perpendicular). See Figure 3.1 for the map with the locations of the profiles.

Profile number	D_{min}-D_{max}	R_{min}-R_{max}	Forested	Orientation
1	0 -2	20-2000	no	perp.
2	0.5-3.5	20-160	no	paral.
3	1-3.5	20-900	at D_{min}	paral.
4	0.5-2.5	20-110	at D_{min}	perp.
5	1-2	20-3000	yes	perp.
6	0.5-3.5	20-260	mixed	paral.
7	1-3	30-80	mixed	perp.
8	1-3	20-90	yes	perp.
9	1-3	20-340	mixed	perp.

4.5 Soil and bedrock leachates

Generally, the concentrations of calcium, magnesium and strontium were higher for the bedrock leachates (L_{BR}) than in the soil leachates (L_S, Table 4.3). Only the calcium concentration in the second-stage hillslope leachate was similar to the bedrock leachates (65-75 mg L^{-1}). Sodium concentrations were comparable in the soil and bedrock leachates for the first leaching stage (L1), and higher in the bedrock leachates for the second leaching stage (average sodium concentration $L2_{BR}$: 3781 μgL^{-1}, $L2_S$: $\sim$ 1000 μgL^{-1}). The concentrations of all other measured elements were higher in the soil leachates than in the bedrock leachates (Table 4.3).

For some solutes, the characteristic groundwater concentrations at each sampling location were also distinctive in the leachate concentrations (shown for iron, zinc, strontium and magnesium in Figure 4.5). For instance, iron concentrations are characteristic for riparian zone groundwater (Type I) and were highest in the riparian soil leachate (L1 riparian soil: 8031 μgL^{-1}, vs. 52.8-864.4 μgL^{-1} for the other leachates). Similarly, zinc, which is characteristic for groundwater Type II, was highest in the hillslope leachate (L2) and strontium, characteristic for groundwater Type III, was highest in the bedrock leachates (both leaching stages). Lastly, the release of magnesium and calcium was particularly high for the silty bedrock in the second leaching stage. This might have contributed to the high magnesium concentrations measured in the groundwater wells close to the outcrop where this bedrock was sampled.

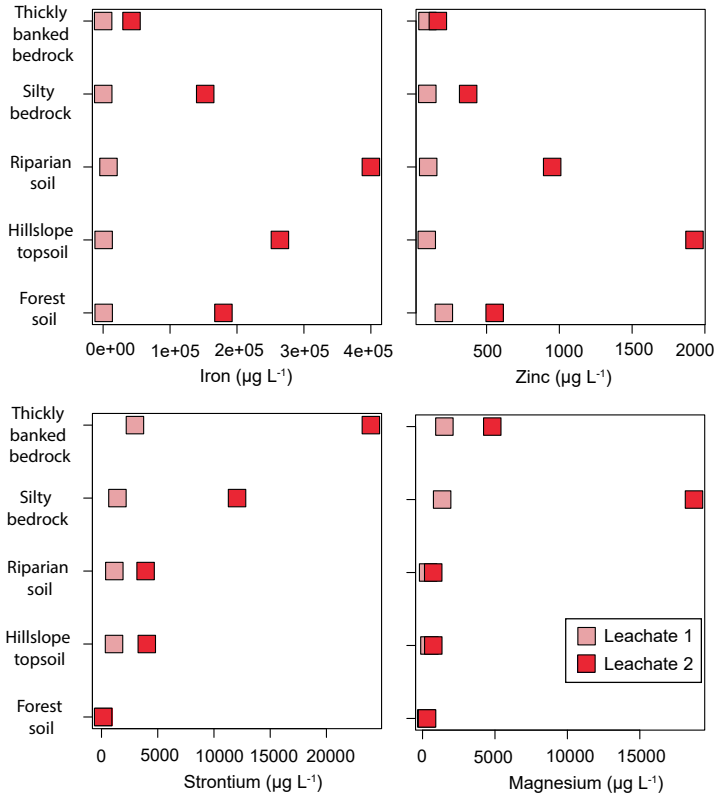

Figure 4.5 – Concentrations of iron, zinc, strontium and magnesium in bedrock and soil leachates after the first (rose squares) and second (red squares) leaching stage. Please note the different x-axes for the different solutes.

Table 4.3 – Bedrock and soil leachate concentrations in μg L^{-1}. See Figure 3.1 for the map of the sampling locations.

First leaching stage (L1)					
Solute	Bedrock	Silty bedrock	Riparian soil	Hillslope soil	Forest soil
Ca	555892.6	474311.1	162061.3	228793.2	27978.6
Sr	2972.3	1415.6	1139.8	1113.1	181.4
Na	2273.6	1628.9	2093.4	1869.1	1327.9
K	15544.7	13510.1	4171.7	5312.8	5616.2
Mg	15083.9	13521.7	3859.1	4739.7	2573.3
Ba	103.3	71.6	207.5	178.7	237.7
Mn	91.0	64.9	16928.7	13698.9	5733.3
Fe	52.8	198.9	8031.4	784.2	864.4
Cu	40.4	<2	20.7	17.9	22.4
Ni	29.4	7.9	24.8	26.8	63.1
Zn	<10	<10	98.8	89.0	206.3
Co	2.2	1.4	13.5	13.2	88.5
Cr	<2	<2	7.7	5.1	7.8

Second leaching stage (L2)					
Solute	Bedrock	Silty bedrock	Riparian soil	Hillslope soil	Forest soil
Ca	7516943.8	6550512.8	396825.3	654247.7	20859.9
Sr	23951.2	12052.2	3921.0	4026.3	145.7
Na	3244.0	4318.4	1032.3	1093.8	880.7
K	5957.9	10737.4	7261.4	8417.7	5037.3
Mg	48128.0	187407.0	7368.1	7419.8	3121.4
Ba	359.9	505.5	6775.7	8696.9	1193.0
Mn	6169.2	6678.2	35811.7	161164.1	17351.7
Fe	42437.5	152520.8	400010.0	264114.0	179752.9
Cu	10.4	51.7	502.8	657.5	307.8
Ni	86.0	121.6	254.8	722.7	97.4
Zn	165.2	373.3	950.6	1928.5	555.4
Co	26.5	194.2	103.3	534.0	220.4
Cr	<2	60.2	46.7	40.5	114.3

Chapter 5

Source areas for streamflow

5.1 Differences in rainwater, soil water and groundwater

The composition of rainwater, soil water and groundwater was distinctly different for various solutes (Table 5.1). Solute concentrations were generally very low in rainfall, except for chloride and sulfate, for which the average rainfall concentration was ∼0.4 and ∼0.15 times the average groundwater concentration, respectively. Rainfall was also characterized by higher, and more variable, D_{ex} values (average D_{ex} in rainfall: 14.1 ‰ vs. 12.0 ‰ for both groundwater and soil water). Solute concentrations in soil water were especially high for barium (SW_{avg}: 37350 $\mu g L^{-1}$; GW_{avg} 99 $\mu g L^{-1}$), potassium and chloride. Groundwater had relatively high concentrations of calcium, sodium, zinc and copper.

Given the characteristic concentrations in different groundwater types (section 4.2), soil water and rain water (Table 5.1), a set of tracers could be identified that reflects contributions from different water sources and landscape elements to stormflow (Figure 5.1). For instance, copper concentrations are high in Type II groundwater (hillslopes and forested areas) whereas they are low in the other water compartments. Increased copper concentrations in stormflow were thus interpreted as larger Type II groundwater contributions. Similarly, higher iron and manganese concentrations could be indicative of larger contributions from riparian areas (Section 4.2), and increasing barium or chloride concentrations might

indicate soil water contributions. For some solutes the interpretation was less straightforward because the concentrations were high in two water compartments. For instance, potassium concentrations are high in soil water but also in Type II groundwater. In addition, potassium has a geogenic origin and it is influenced by biological processes. Using potassium as a tracer is thus less certain.

As mentioned in section 3.5.3, most tracers were not conservative. As such, any interpretation regarding source water contributions inferred from the streamwater solute concentrations should be taken as a rough estimation, rather than an exact determination. Additionally, the variability of various tracer concentrations in different water compartments was large, which adds to the uncertainty. All the above mentioned tracers were used in addition to the isotopic compositions.

Table 5.1 – The average concentrations (± standard deviation) for all groundwater (GW_{avg}; n=335), soil water (SW_{avg}; n=116), and rainfall (P_{avg}; n=156) samples.

Solute	Unit	GW_{avg}	SW_{avg}	P_{avg}
D_{ex}	‰	12.0±0.8	12.0±2.4	14.1±3.2
$\delta^{18}O$	‰	-11.0±0.9	-10.4±1.6	-12.3±4.0
$\delta^{2}H$	‰	-76.0±7.5	-70.8±12.4	-84.4±33.0
Ca	μgL^{-1}	56993.7±21966.1	22261.7±27287.8	213.4±202.7
SO_4	μgL^{-1}	3600.0±5112.5	1602.0±3061.9	623.1±980.1
Mg	μgL^{-1}	2235.7±1730.3	13612.8±10924	26.6±18.9
Na	μgL^{-1}	1587.6±2672.7	839.1±565.0	148.7±153.5
Cl	μgL^{-1}	830.8±1076.5	1070.3±1026.6	327.1±348.7
Zn	μgL^{-1}	593.9±1745.7	23.3±12.5	19.3±43.0
Mn	μgL^{-1}	592.4±1111.6	139.9±326.2	1.3±1.4
K	μgL^{-1}	530.1±428.0	754.1±970.8	92.2±91.9
Fe	μgL^{-1}	390.7±1271.1	254.3±775.9	3.5±7.1
Ba	μgL^{-1}	99.2±171.6	37350±27637	4.8±11.8
Cu	μgL^{-1}	64.9±143.7	5.2±9.0	1.4±1.0
Ni	μgL^{-1}	3.2±4.1	2.5±1.5	0.3±0.3
Co	μgL^{-1}	0.8±1.05	0.9±1.1	0.02±0.02
Cd	μgL^{-1}	0.05±0.08	0.03±0.06	0.1±0.2

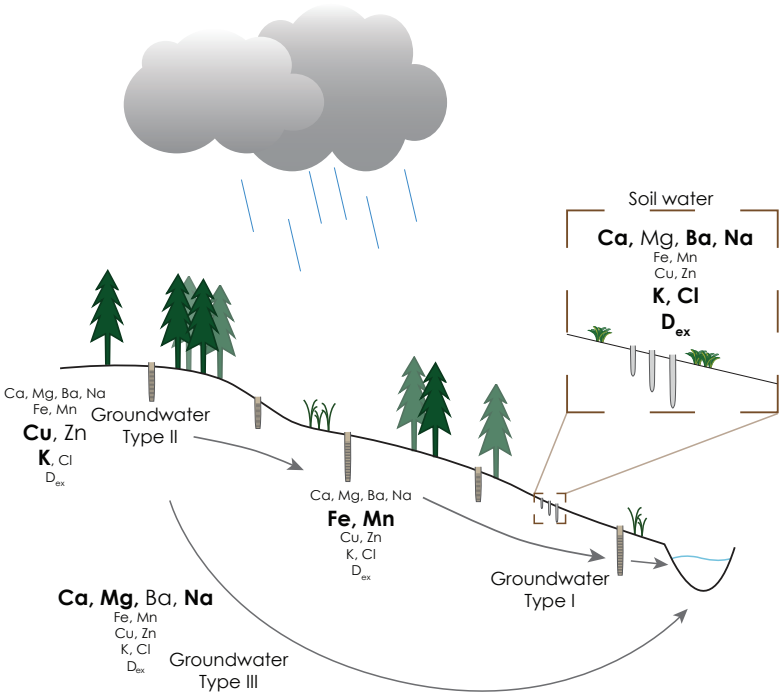

Figure 5.1 – Schematic representation of a hillslope cross-section showing the different (ground)water compartments, the tracers used in combination with δ^2H and δ^{18}O to characterize source areas, with the most important tracers for each water compartment in bold. For most elements, the concentrations were low in rainfall compared to the concentrations in the other water compartments. High potassium, barium and chloride concentrations and high deuterium excess are indicative of soil water. For shallow groundwater, the concentrations of copper and potassium were higher at (forested) ridge locations; for sites with a water tables that is persistently close to the surface, the concentrations of iron and manganese were higher. Higher concentrations of geogenic solutes (calcium, magnesium and sodium) indicate longer residence times. The isotopic composition for the different water compartments depends on the composition of recent and current precipitation.

5.2 Source areas for baseflow

The streamwater samples plotted near the groundwater samples in the principal component space, but were not the average of all groundwater samples (PC1 and PC2 are shown for SC3 and SC6 in Figure 5.2). The streamwater samples were most similar to the median coordinates of groundwater Types I and II (i.e. the riparian-like and hillslope groundwater, respectively) at all sites except for C1. During SC1 and SC6, which took place shortly after snowmelt, streamwater was most similar to groundwater Type II, whereas it was more similar to groundwater Type I during the other campaigns. This suggests that groundwater contributions from these landscape elements were relatively more important than groundwater Types III and IV. The streamwater at catchment C1 was most similar to groundwater Type IV (in SC1 and SC6) and groundwater Type III (in all other campaigns).

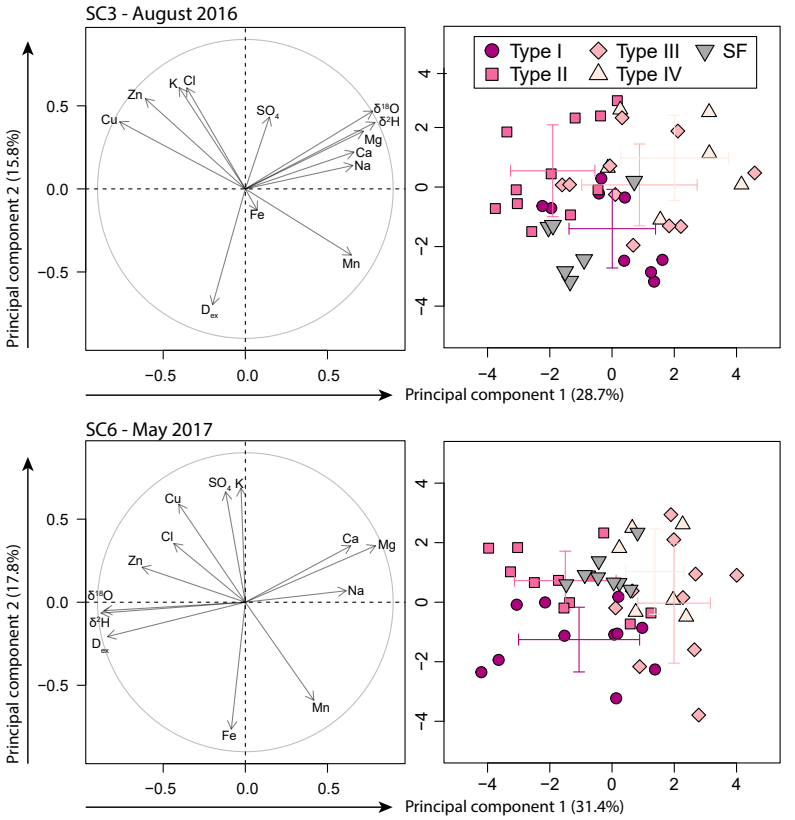

Figure 5.2 – Two-dimensional PCA results for groundwater chemistry during SC3 (top) and SC6 (bottom). The biplots (left) show the correlation of the element concentrations to the principal component axes; the length of the arrow represents the strength of the correlation. The total explanatory power of the individual principal components is indicated on the axes. The mixing plots (right) show the projection of the streamflow (SF, grey reversed triangles) and all groundwater samples colored by groundwater type based on the clustering on the M_{RD} values (Section 4.2; dark purple circles = Type I, pink squares = Type II, rose diamonds = Type III, light rose triangles = Type IV) in the same two-dimensional principal component space, as well as the average and standard deviation/error for each groundwater type (error bars). SC6 is representative for early summer (after-melt) conditions, while SC3 is representative for typical summer conditions. Figure from paper I.

5.3 Source areas for stormflow

Two-component hydrograph separation indicated that most stormflow was 'old' water (Figure 5.3, Table 5.2). The maximum event-average event water fraction (f_e) was largest for event II (f_e = 0.24±0.31) and smallest for event I (f_e = 0.09±0.16). The highest event water fraction for event II occurred when the connected area was relatively small (Table 3.2, see Paper II for a detailed description). The event-water fractions were higher for event III than the similarly sized event IV. This might be the result of the much smaller hydrologically connected area and relatively high peak rainfall intensity (I_{p-max}: 24 mm h^{-1} vs 10 mm h^{-1} for event IV, Table 3.2).

The explanatory power of the first two principal components, including all stormflow, soil water and groundwater samples, was 48.4% for event I (PC1: 29.8 %; PC2: 18.6 %, representative for a small event) and 51.6% for event III (PC1: 32.8%; PC2: 18.8%, representative for an intermediately sized event; Figure 5.4). The principal component axes were most strongly determined by the calcium concentrations and EC (orientation close to PC1 for both events), the isotopic composition (more so in event III) and to a lesser extent by concentrations of copper, magnesium, and potassium, and D_{ex}. It was possible to calculate the relative fractions of groundwater, soil water and rainwater in stormflow for all events based on EMMA, but the uncertainties were very large (Table 5.2). The groundwater fractions (f_{GW}) were dominant for all events (range f_{GW}: 0.46±0.50 to 0.31±0.19). Event-average soil water fractions were negligible (f_{SW} ~0) during events I, III and IV, and considerable (0.24) during event II (Table 5.2). The event-average pre-event water fractions based on the EMMA (i.e., the sum of the groundwater and soil water fractions) was similar to the pre-event water fraction estimated using δ^2H as a tracer in the two-component hydrograph separations (range f_{GW} + f_{SW}: 0.70 to 0.81 vs range f_{pe}: 0.76 to 0.91; equation 3.2).

However, the most striking aspect of the mixing plots was the small change in the composition of stormflow compared to the variation in end-

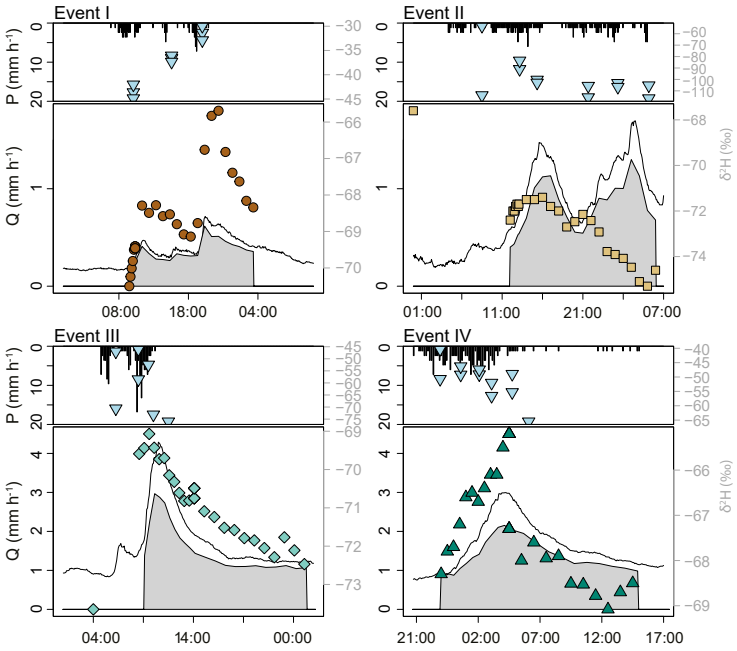

Figure 5.3 – Hydrographs and hyetographs for the four events (I-IV). For each event, the upper panel shows rainfall at 10-min resolution (mm h^{-1}, bar graph) and the isotopic composition of the rainfall (δ^2H in ‰, light blue reversed triangles), while the lower panel shows the discharge at the catchment outlet (mm h^{-1}, solid line), the isotopic composition of streamflow (δ^2H in ‰; brown dots, light brown squares, turquoise diamonds and green triangles for event I-IV, respectively), and the pre-event water fraction in streamflow based on two-component hydrograph separation using δ^2H as a tracer for the period that was used to calculate the event-average pre-event water fractions (grey polygon). Please note the differences in scale for discharge. Figure from paper II.

members (Figure 5.4, Table 5.1). The observed changes in solute concentrations in streamflow were largest during event III (e.g., changes of 25 μg L^{-1} for Ba, 39 mg L^{-1} for Ca, and 5 ‰ for δ^2H) but this change was similar to the standard deviation of all the groundwater samples (44 μgL^{-1} for Ba, 27 mgL^{-1} for Ca and 5.9 ‰ for δ^2H). This large variability induced considerable uncertainties in the calculated fractions, and inhibits strong conclusions on the source areas.

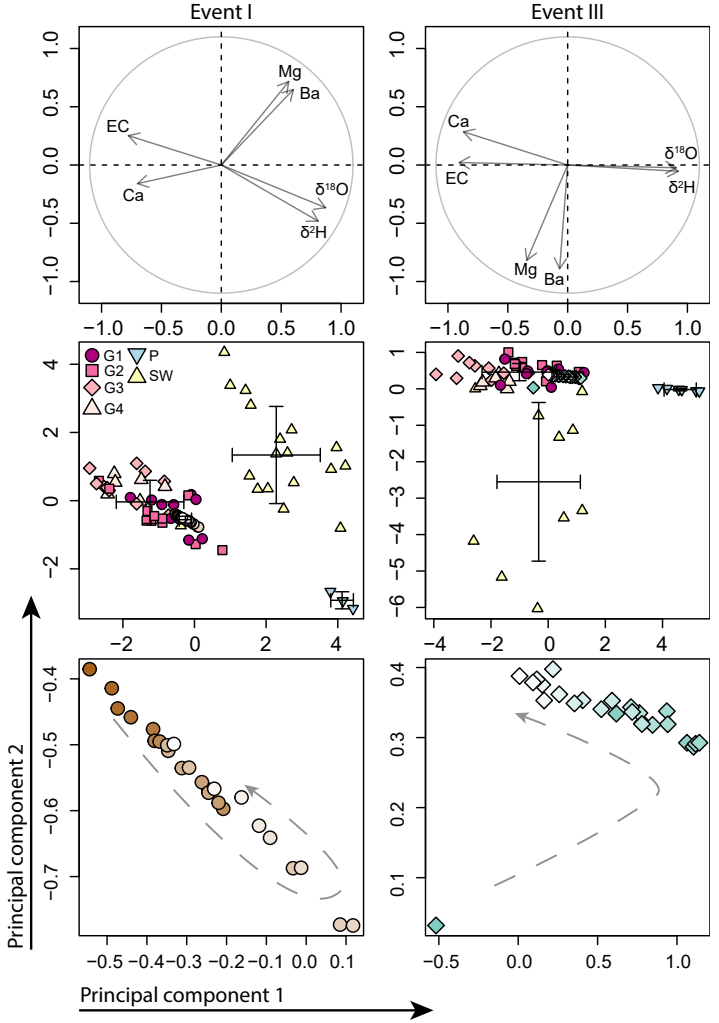

Figure 5.4 – Mixing analysis results for event I (left) and event III (right). Event I is representative of a small event, whereas event III is representative of an intermediately sized event. The length of the arrow in the biplots (top) represents the explanatory power. The mixing diagrams based on the first two principal components (middle) show the rainfall (P, light blue reversed triangles), soil water (SW, yellow triangles), and groundwater samples (purple circles, pink squares, rose diamonds and light rose triangles, representing groundwater types I-IV; section 4.2). The streamflow samples (SF) are indicated in brown (start event I) and turquoise (start event III) and fade towards white (end of event). The average and standard deviation for each component are shown with error bars. The bottom row shows a zoom for the streamflow samples only, in which the direction of change is indicated with the grey arrow (dashed lines). Figure from paper II.

Table 5.2 – Event-average fractions of groundwater (f_{GW}), soil water (f_{SW}), and rain water (f_P), determined by three component EMMA, and the pre-event water fraction as determined by two-component hydrograph separation based on δ^2H (f_{pe}), and the corresponding uncertainties.

Event	f_{GW}	f_{SW}	f_P	uncertainty	f_{pe}	uncertainty
I	0.81	~0	0.19	0.16	0.86	0.28
II	0.49	0.27	0.24	0.14	0.76	0.61
III	0.72	0.01	0.27	0.16	0.81	0.69
IV	0.74	0.01	0.25	0.14	0.78	0.25

The concentrations estimated based on the assumption of conservative mixing between rainfall and baseflow (C_{es}, Equation 3.3) differed from the measured stormflow concentrations (C_Q) for almost all solutes (Figure 5.5). The measured concentrations for geogenic solutes (shown for calcium in Figure 5.5a) were lower than estimated. This could be due to mixing with a source with lower calcium concentrations (for instance soil water, or riparian groundwater; Table 5.1). The measured concentrations of sulfate were lower than estimated for events I, III and IV. For potassium concentrations there was no clear pattern: the concentrations were underestimated and overestimated at both lower and higher discharges, which is probably due to the high variation in potassium concentrations in soil water and groundwater (Table 5.1). The measured concentrations of cobalt, copper and iron were slightly lower than the estimated concentrations for low discharge but (much) higher during high discharge (Figure 5.5). For copper and nickel this could be due to hillslope contributions, whereas for iron and cobalt it could be due to increased contributions from the riparian areas. There was no distinct threshold in the relation between discharge and C_Q/C_{es}.

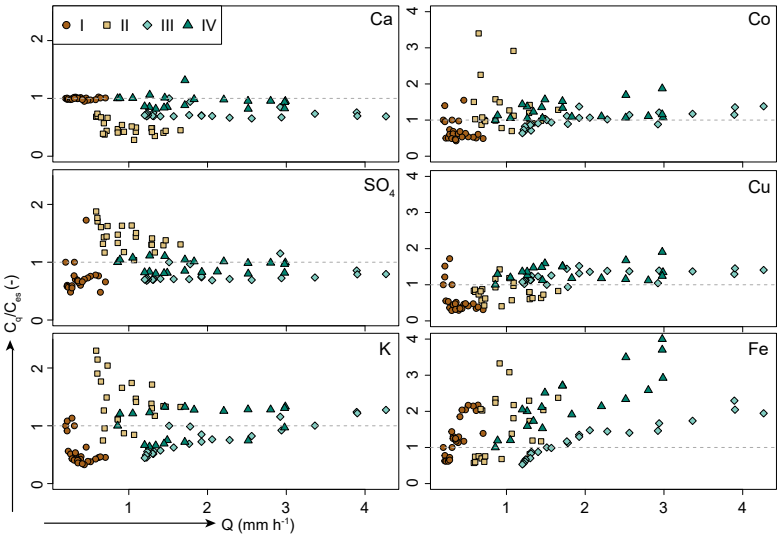

Figure 5.5 – The ratio of the measured (C_Q) and estimated stormflow concentrations (C_{es}; Equation 3.3) for calcium, sulfate, potassium, cobalt, copper, and iron as a function of the specific discharge (Q) at the catchment outlet. The dashed grey line indicates where C_Q and C_{es} are equal; the different symbols reflect the different events (I-IV). Note the difference in scale for the left and right column. Figure from paper II.

Chapter 6

Uncertainty in isotope hydrograph separation analyses

A baseflow sample is often used to characterize the pre-event water composition in hydrograph separation analyses. However, given the large variability in the shallow groundwater isotopic composition (Section 4.1) and changes in source areas to streamflow during stormflow conditions (Chapter 5), this might result in large uncertainties in the estimated pre-event water fraction. Furthermore, these uncertainties might go unnoticed if the groundwater variability is not characterized. The following sections describe the effects of spatial variability in the shallow groundwater isotopic composition on pre-event water characterization and hydrograph separation results, and include an estimate of how many wells should be sampled to capture the spatial variability in the pre-event water composition.

6.1 Sensitivity of hydrograph separation results

Collecting only one groundwater sample to characterize C_{pe} can result in a large range of pre-event water fractions. For this dataset, the maximum differences in f_{pe} ranged from 0.28 to 0.47 for event I and event IV, respectively (Figure 6.1).

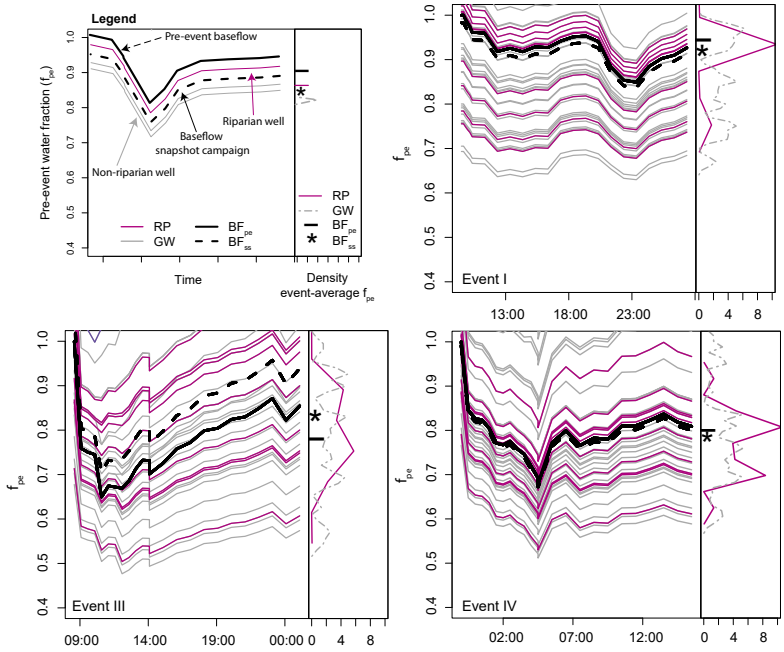

Figure 6.1 – Time series of the calculated pre-event water fraction (f_{pe}) for events I, III and IV using δ^2H as a tracer and either the pre-event baseflow sample (BF_{pe}, solid black line), the baseflow sample taken during the snapshot sampling campaigns (BF_{ss}, dashed black lines), a sample from each riparian well (RP, purple lines), and all other wells (GW, grey lines) to represent the isotopic composition of the pre-event water (C_{pe}), as well as the frequency distribution of the event-average pre-event water fraction (Kernel density plot, right side of each subplot) for each method used to represent the pre-event water composition (BF_{pe}: black dash, BF_{ss}: black asterisk, RP: purple solid line, GW: grey dashed line). See section 3.5.5 for a detailed explanation of the different pre-event water characterization methods. Figure from paper III.

The event-average pre-event water fractions were also sensitive to the choice of pre-event water characterization (Figure 6.2). The spread in the event-average f_{pe} was, not surprisingly, largest for the ensemble of the calculations based on one groundwater sample, as they spanned the whole range of possible isotopic compositions from which the average groundwater composition was calculated. Selecting only riparian wells to characterize C_{pe} resulted in the highest event-average pre-event water

fraction compared to a selection from all groundwater wells or a pre-event baseflow sample (Figure 6.2). However, ultimately this depends on the relative composition of each water compartment, and can thus differ for different events.

The median uncertainties in the event-average pre-event water fraction (W_{fpe} in equation 3.4) ranged from a low of 0.04 when using pre-event baseflow to characterize the pre-event water composition for event I (BF_{pe}) to a high of 0.92 (median value for all combinations) when using three riparian wells to characterize the pre-event water composition (RP3) for event III. Not surprisingly, the calculated uncertainties in f_{pe} were smallest for events in which the pre-event water composition and rainwater composition were more different.

Increasing the number of samples to determine C_{pe} reduced the sensitivity and the uncertainty in the pre-event water fraction (Figure 6.2; Equation 3.4). W_{fpe} was largest when three samples were used to calculate the pre-event water composition due to the high t-value for small sample sizes and the high standard deviation for some of the combinations of samples (see right column of Figure 6.2). As a result, the reduction in the median uncertainty was largest when the number of samples increased from three to six samples. Moreover, W_{fpe} was smallest for the calculations based on a baseflow sample or one groundwater sample because the uncertainty of the pre-event water composition (W_{Cpe} in Equation 3.4) was assumed to be equal to the measurement precision of the isotope analyzer for this situation. For the uncertainty estimation for the pre-event water fraction based on the selection of three, six or nine samples or the average composition of all (riparian) groundwater samples, the uncertainty was based on the standard deviation of the selected samples and corresponding t-value for small sample sizes, and thus, to some extent, reflects the variability in the pre-event water composition.

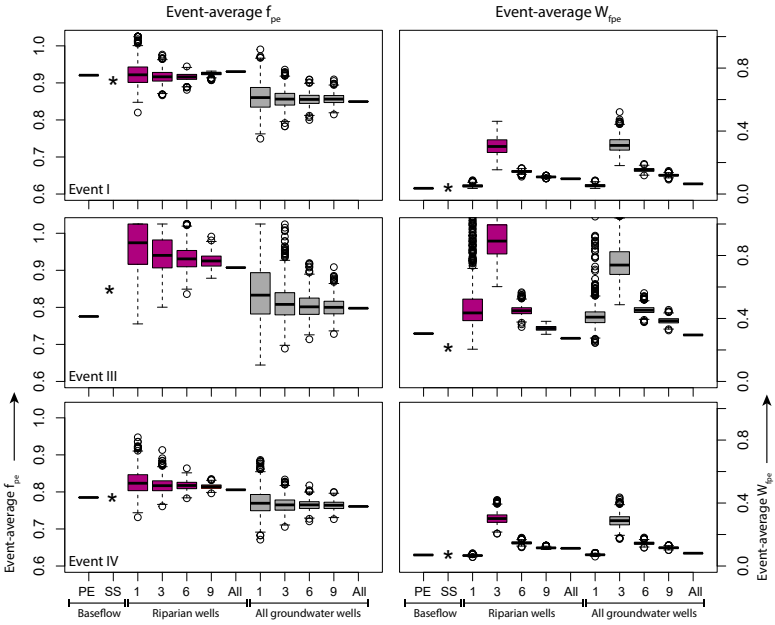

Figure 6.2 – Boxplots of the event-average pre-event water fractions (f_{pe}, left) and the associated uncertainty (W_{fpe}, right) for events I, III and IV (rows), when the pre-event water composition is represented by a baseflow sample taken before the event (PE, dash), a few days later during the snapshot campaign (SS, asterisk), and the average isotopic composition based on samples from one, three, six or nine randomly selected wells in riparian-like areas (purple), or across the entire catchment (grey), and based on the average composition of all riparian wells (All, purple, n=11) and all wells across the catchment (All, grey, n=38). All boxplots are based on 1000 random selections of wells. Figure from paper III.

6.2 Number of samples required to characterize pre-event water

During baseflow campaigns SC4 and SC9, which closely followed the studied events, the spatial variability in the isotopic composition of the shallow groundwater was large. δ^2H varied from -86.3 to -67.8 ‰ and from -80.9 to -57.2 ‰ , respectively. Riparian groundwater was slightly more enriched than the average of all sampled groundwater wells for both snapshot campaigns (Figure 6.3). This difference was larger than twice the laboratory precision (0.6 ‰ δ^2H), but statistically not significant.

The difference between the 5^{th} to 95^{th} percentile of the average composition for six randomly selected groundwater samples was 6.3 and 4.6 ‰ for SC4 and SC9, respectively (Figure 6.3). For nine randomly selected samples, the composition ranged between 5 and 3.6 ‰, respectively. The difference between the 5^{th} and 95^{th} percentile of the calculated average groundwater composition was less than 2.5 ‰ (i.e., half of the average change in the isotopic composition of streamwater during the studied events) when more than 21 respective 16 random samples were used to determine the average composition.

The calculated variability of the groundwater isotopic composition also decreased with increasing sample size (Figure 6.3). The 5^{th} to 95^{th} percentile of the standard deviation of the isotopic composition of the groundwater for six randomly selected groundwater samples were 2.7 and 7.7 ‰ for SC4 and 1.9 and 5.5 ‰ for SC9. For nine randomly selected samples, they were 3.2 and 7.1 ‰ for SC4 and 2.5 and 5.1 ‰ for SC9. The difference between the 5^{th} and 95^{th} percentile of the calculated standard deviation of the groundwater was less than 1.2 ‰ (which equals twice the laboratory accuracy), as soon as more than 29 respective 22 random samples were used to determine the variability in δ^2H of groundwater.

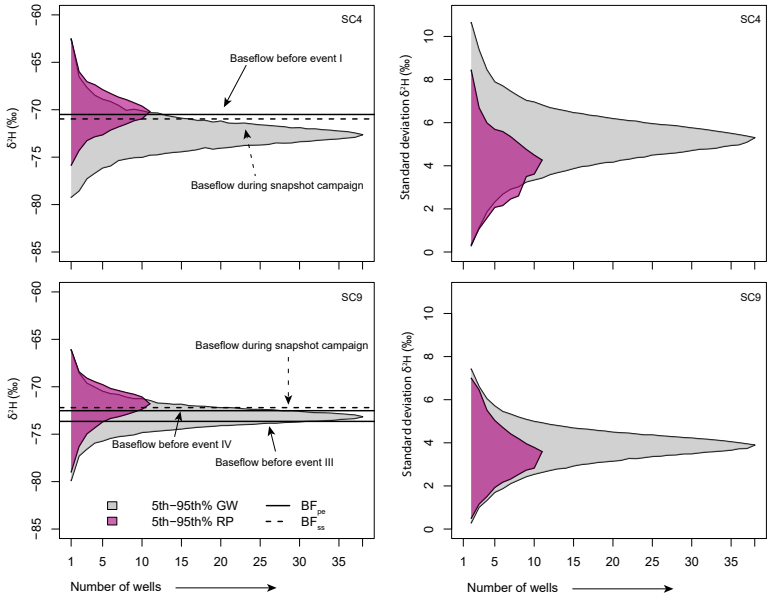

Figure 6.3 – The 5^{th} to 95^{th} percentile of the average (left) and standard deviation (right) of the isotopic composition (δ^2H) of groundwater as a function of the number of randomly selected groundwater samples (n=1 to 38 for all groundwater wells (GW, grey) and n=1 to 11 for the riparian wells (RP, purple), 1000 repetitions) based on samples taken during SC4 (upper panels) and SC9 (lower panels). The horizontal lines indicate the isotopic composition of baseflow at the outlet during the snapshot campaign (BF_{ss}, dashed lines), and prior to the event (BF_{pe}, solid line). Figure from paper III.

Chapter 7

Discussion

The results presented in this thesis demonstrate that the spatial variability in shallow groundwater isotopic and chemical composition in a small pre-alpine headwater catchment can be very large. They indicate the importance of landscape elements and aquifer chemical composition for the spatial distribution and composition of groundwater, and show that this variation is temporally persistent so that it can be used to identify source areas to streamflow. Furthermore, this work shows that baseflow might not be representative for all water that is stored in the catchment, which contrasts the common approach in many hydrograph separation studies in small headwater catchments. The effects of spatial variability in groundwater compositions on source-area analyses can be large. This should motivate us to increase our typical sampling efforts, so that our analyses yield robust results.

7.1 Spatial variability in groundwater chemistry

The nine baseflow snapshot campaigns in 2016 and 2017 highlight the large spatial variability in shallow groundwater chemistry across the 20-ha Studibach catchment during baseflow conditions. In this analysis, it is assumed that concentrations measured during individual campaigns are typical for that time of year. However, any short-lived extreme

concentrations might have been included in the 'monthly' value, since the wells were sampled only once per month. Several other studies have characterized spatial variations in groundwater composition. However, they focused on the isotopic composition (Carey and Quinton, 2005; Kendall et al., 2001; Klaus et al., 2015a) or solute concentrations (Burns et al., 1998; Walker et al., 2003; Welsch et al., 2001), rather than considering both. What these studies have in common is that they all report large spatial variabilities (e.g., between 0.28 ‰ δ^{18}O for six groundwater wells in northern Canada (Carey and Quinton, 2005) to 4 ‰ δ^{18}O for the artifical hydrohill catchment (Kendall et al., 2001)). The variabilities are attributed to heterogeneities in subsurface flow paths and velocities, for instance due to macropore flow. In our case, the highly heterogeneous Flysch bedrock material may have additionally contributed to the large variability in groundwater composition in the Studibach cathment.

For the Studibach, the spatial variability was larger than the temporal variability for most solutes, so that four shallow groundwater types could be distinguished based on the relative differences from the catchment average. Such classifications can be useful to study runoff processes and catchment scale connectivity, as has been shown for catchments in Japan (Sidle et al., 2000) and New Zealand (McGlynn and Seibert, 2003). The classification relied strongly on differences in concentrations of transition metals, and thus likely reflect the dominant redox conditions. Since these elements are sensitive to reactive transport processes, care should be taken with interpretations of source-area analyses based on these tracers, because transport and mixing likely changes oxygen concentrations in the groundwater (a dominant control on redox states; Kaushal et al., 2018). Fortunately, the short response times in this steep catchment constrain reactive processes at least to some degree.

The mainly forested upper hillslope and ridge sites were characterized by high concentrations of copper, zinc and lead, which might be (partially) due to higher soil acidity and subsequent leaching from the forest soils (Harter, 1983), or by larger atmospheric inputs compared to the non-forested parts of the catchment (Shotyk, 1996; Shotyk et al., 2000). Sites

with persistent shallow groundwater levels were characterized by high concentrations of mangenese and iron, which is likely the effect of almost permanent reducing conditions (Megonigal et al., 1993). Adsorption of heavy metals (such as zinc, copper or lead) on manganese and iron oxides (McKenzie, 1980) can consequentially have led to the contrasting concentrations between the hillslopes and riparian areas. Both groundwater types had slightly enriched isotopic compositions compared to the other two groundwater types, which indicates that they might have been fed by more recent precipitation (as sampling occurred in the snow-free season). No evidence for evaporative enrichment was found in any of the groundwater samples.

Previous studies showed that nitrate (Welsch et al., 2001) and base cation Burns et al. (1998) concentrations could be explained by local differences in flushing frequency and bedrock drainage area. For the Studibach, the correlations of solute concentrations to individual site attributes were mostly weak. This relation might have been weakened by the vertical variability in groundwater concentrations, which was not investigated. Groundwater can be layered, even within a shallow aquifer, and the degree of layering can differ on small distances. For instance, Martin et al. (2004) showed that for two adjacent catchments, the layering of the groundwater nitrate concentrations was much more pronounced in one catchment than in the other. For the data collection of this work, all wells were purged before each snapshot campaign such that they were supposedly a mixture of all water in the well profile. However, since the contribution from each depth depends on the permeability and saturation level of the groundwater layer at that depth, the relative contribution from different groundwater layers may have been different during the different campaigns. This may have weakened the spatial pattern in groundwater compositions. Conversely, the weak relation might also just be the effect of high variability in the upper part of the shallow aquifer. Legout et al. (2007) showed that the variability in solute concentrations decreased with depth for the Kerbernez catchment in France, in which the groundwater depth extends to 15-20 m. They also showed that solute concentrations in the upper 1-1.5 m (which equals the depth of most of the Studibach wells) were extremely variable.

Although the measures of surface and subsurface topography and aquifer chemical composition were not sufficient to explain the variable solute concentrations, the geophysical profiles did (qualitatively) indicate that the interplay of subsurface storage and surface topography determines the groundwater types (Section 4.4). Furthermore, the rugged bedrock topography that was captured in some of the profiles showed that isolated groundwater pockets might exist (cf. Brammer and McDonnell, 1996), which likely increases the spatial variability in the groundwater. Also the chemical composition of the surrounding material was shown to be important, as the soil and bedrock leachates indicated that some of the contrasting element concentrations in the groundwater types were similarly contrasting in the leachates from these respective areas. However, these results are statistically not significant, as only one sample was available per landscape element. Interpretations should thus be made with care. These findings demonstrate the importance of knowledge on bedrock and soil properties, and should motivate catchment scientists to gather more information on subsurface geometry and chemistry, and incorporate this in hydrological analyses.

7.2 Source areas for streamflow

The changes in streamwater composition during events and seasonally indicate that source areas to the stream are dynamic. Baseflow was never an equal mixture of the different groundwater types, which indicates that baseflow does not reflect the average composition of the groundwater in the catchment (Figure 5.2). During the majority of the campaigns the streamwater chemistry most strongly resembled groundwater with a riparian-like chemistry (Type I), except at C1. For the campaigns that occurred shortly after the snowmelt period the highest similarity was to the "hillslope" groundwater (Type II). This might reflect hillslope-riparian zone-stream connectivity (Jencso et al., 2009; von Freyberg et al., 2014) as discharge was six to ten times higher during the campaigns shortly after snowmelt than during the other seven campaigns (Table 3.1), and the catchment was thus significantly wetter. The increased connectivity

during wet conditions is also reflected in the stronger correlation between topography and water levels during this period (Rinderer et al., 2014). The similarity of streamflow at site C1 to the "deeper groundwater" (Type III) confirms the importance of deeper groundwater flow to sustain stream baseflow that is often applied in conceptual models (Sklash et al., 1976) or large scale studies (Tague et al., 2008).

Simple mixing of baseflow and rainfall could not explain the solute concentrations in streamwater during events. However, the differences between the measured and expected concentrations with increasing discharge showed consistent patterns for different solutes (e.g., iron, copper and strontium). Since these tracers are not conservative, these data should be carefully interpreted. Hence, we can only say that the differences are likely to be, at least partly, the result of contributions from source areas that did not contribute to baseflow. This is especially true at peak flow and during discharge recession, when the hydrologically connected area is increased with respect to baseflow conditions.

Simulations of the active and connected area confirm that large changes occur in the hydrologically connected area, and that near-stream areas are most often connected and respond first to rainfall. This shows their importance for the rapid generation of streamflow (Rinderer et al., 2019; van Meerveld et al., 2019). The simulations of the hydrologically connected area for the events described in this thesis (see detailed results in Paper II) also show that connectivity can change little during long, low-intensity rainfall events, and that large areas can remain disconnected, even during peakflow. As such, it is important to characterize the variability in areas that are likely to contribute to stormflow when performing source-area analyses, and to consider the variability in the entire catchment when studying extreme events. As such, the uncertainty estimations presented in this dissertation might be overestimating the uncertainty during small or intermediate events, because all wells were included, not only the wells that were likely to contribute to streamflow. However, selecting only riparian wells resulted in very similar uncertainty ranges, and thus would not have changed the interpretation of the results.

Despite large changes in the hydrologically connected area and the large spatial variability in groundwater composition, there was no distinct threshold in the deviation of stream chemistry from simple conservative mixing of rainfall and baseflow (Figure 5.5). The gradual change in chemistry might reflect the gradual increase in connectivity, and mixing of the contributions from many small landscape elements, such as described by Abbott et al. (2018). The hydrochemical signal might have also been obscured by heterogeneities in subsurface mixing processes, as for instance reported for mixing of vadose zone water and groundwater in the capillary fringe during recharge events (Silliman et al., 2002). Such heterogeneities can lead to groundwater layering, and since hydraulic conductivities decrease with depth, the groundwater contribution to the stream will not be the integrated signal from the entire saturated profile but rather reflect the upper part of the profile.

The results from the stormflow sampling did show that the changes in streamflow composition during four rainfall events were much smaller than the spatial variability in groundwater and soil water composition (Figure 5.4). This resulted in considerable uncertainties (range W_{fpe}: 0.14 - 0.16), which inhibited the quantification of contributions from different (groundwater) sources. The results also suggested that soil water fractions were very small ($\sim$0) for three of the four events (Table 5.2).

7.3 Uncertainty in hydrograph separation analyses

The uncertainty of the calculated (pre-)event water fractions depends on the uncertainty of each end-member (Equation 3.4). Therefore, including a realistic estimate of the variability within each water compartment is necessary for robust analyses. The uncertainty estimates presented in this thesis were comparable to or larger than the uncertainties described for other small catchment studies (e.g., Fischer et al., 2017; Pellerin et al., 2008; Penna et al., 2017), and show that uncertainties are likely

large if you properly account for the variability in (pre-)event water composition. Uncertainties in hydrograph separations due to the spatial variability in event-water can be just as large as presented in this thesis for pre-event water variability. For instance, Cayuela et al. (2019) reported an uncertainty of 0.01 – 0.14 for the Can Vila catchment (Spain), and f_{pe} estimations of Lyon et al. (2009) and Fischer et al. (2017) differed more than 50% between computations based on different rainfall sampling locations for the Upper Sabino catchment (Arizona, USA) and Zwäckentobel (Switzerland), respectively. Altogether, these findings suggest that uncertainties in the pre-event and event water fractions can be large, even when the variability that is included in the calculations might still be smaller than the actual variability in the pre-event water composition.

Although many studies use pre-event baseflow to characterize the pre-event water composition (Penna and van Meerveld, 2019), this work shows that pre-event baseflow did not reflect the variability within the catchment, even if it had the same composition as the average groundwater composition. Therefore, when using a baseflow sample to characterize the pre-event water composition the total uncertainty is likely underestimated. Additionally, one might risk misinterpreting runoff generation processes if the composition in transiently connected areas is significantly different than in the areas that contribute to baseflow (see example in Paper III).

This work could demonstrate that the spatial variation within different source areas is large compared to the temporal variation, because the spatial variability was characterized extensively. In other small catchment studies this comparison is often not possible, because of insufficient spatial sampling (Penna and van Meerveld, 2019). Although the variability in groundwater composition can be large, it is not necessary to have a sampling network of $>$30 wells in all research areas. Sampling a limited number of wells can already give an estimate of the variability within the groundwater, and thereby aid a careful interpretation of hydrograph separation analyses. The results indicated that increasing the number of samples from three to six resulted in the largest uncertainty reduction,

and that after sampling six groundwater wells it is probably also clear if the spatial variability is large or if it is not. However, it is concerning that most studies in which the spatial variability was quantified were geographically concentrated to a few regions in the world (five of the six studies mentioned in section 7.1 were conducted in North-America). Hence, in order to find out if the spatial variation is also larger than the temporal variation in other locations (or not) and if the uncertainties are extreme or typical, it is paramount to quantify the spatial variation by sampling groundwater and soil water at multiple sites in more research areas, specifically in different climates and geological settings.

7.4 Conceptual model of the Studibach

The combined findings of this dissertation and earlier studies in the Studibach allow me to refine the conceptual model for the Studibach (Figure 7.1). This conceptual model highlights the importance of the steep slopes and low-permeability soils, and the overall wet conditions that are prevalent in this small headwater catchment for its hydrologic functioning.

Specific discharge increases with catchment size (van Meerveld et al., 2018), which suggests that the lower part of the catchment, where water accumulates, contributes more to streamflow than the upper part of the catchment, from which water is drained. This is reflected for instance in the persistence of groundwater levels in the lower part of the catchment during the 2018 dry spell, while the springs and groundwater wells in the upper part of the catchment fell dry. This is contrary to the findings of Fischer et al. (2015), who suggested that the upper spring zone is most important for baseflow generation, based on a study using calcium, DOC and δ^2H as tracers. This highlights that source area analyses can be affected by the chosen set of tracers.

In most locations, there is an almost permanent water table in the low-permeability gleysols, because these areas are fed by steep, upslope areas, and water is released slowly due to the low permeability of the soil. Flatter, poorly drained areas are found throughout the catchment and occur at various elevations and distances to the stream; their exact location is thus important for connectivity to the stream (Figure 7.1). Seepage to the underlying bedrock is assumed to be limited, as otherwise there would not be a permanent water table above the bedrock in such a large part of the catchment. However, no seepage rates or bedrock permeability measurements are available, and some seepage and exfiltration is needed to explain the groundwater with the deeper flow pathway (Type III) and the mentioned increase in discharge with scale.

The near-stream areas are most important for sustaining stream baseflow,

whereas wetter sites farther from the stream can become important contributors during wetter periods. Groundwater levels respond almost instantaneously to precipitation (i.e., after 1-5 mm, Rinderer et al., 2014), and significant lateral flows can occur when the water table rises into the near-surface layers, where the hydraulic conductivity is much larger (cf. Schneider et al., 2014). Hillslopes and ridges (Type II) are important for streamflow generation during very wet conditions. The slopes are steeper and soils are more permeable, and thus overall less effective at retaining water. During very wet conditions, such as after the spring snowmelt, connectivity between the hillslopes and stream can be established, and hillslopes can even be the largest contributor to baseflow (Figure 5.2). Hillslopes modulate hydrologic connectivity between upslope areas and the stream. Rinderer et al. (2014) suggested that after as little as ∼5-10 mm of rainfall the hillslope groundwater levels rise, and hillslope groundwater can potentially contribute to streamflow. When this occurs, any isolated topographic depressions (i.e., not adjacent to the stream) can become connected to the stream. The expansion of the hydrologically connected area is rapid, and occurs simultaneously in many landscape elements across the catchment, but many areas also remain disconnected from the stream for extended periods of time (Rinderer et al., 2019). This leads to gradual changes in streamwater composition at the catchment outlet, and challenges the quantification of connectivity and contributions from different source areas using hydrochemical tracers.

Stream responses to rainfall are quick, and pre-event water fractions are high. This suggests that overall storage deficits within the catchment are small, at least in the areas that quickly become connected to the stream. The large pre-event water fractions also suggest that surface runoff is relatively unimportant, although it occurs frequently, especially at bareland sites and meadows (van Meerveld et al., 2018). As such, it seems likely that surface runoff infiltrates at more permeable locations before reaching the stream. The role of water stored in the unsaturated zone remains unclear. Soil water contributions were minor for most of the events described in this thesis, but considerable for one event. The relative soil water contribution did not seem to depend on event characteristics (e.g.,

stage height, connected area, rainfall intensity). Hence, the importance of soil water might be controlled by factors that are not yet captured, such as the high variability in soil water amount in this complex terrain.

Also groundwater storage is highly variable in the Studibach. Apart from water storage in riparian-like or wetland areas, the geophysical data suggest that water might also be stored in groundwater pockets at the soil-bedrock interface (Figure 4.4). These groundwater pockets can be activated as recharge occurs, but also cause spatial variability in the groundwater levels, and thereby control subsurface hydrological connectivity. The geophysical profiles also reveal that the resistivities below the interpreted bedrock interface are not low at all locations. This could indicate that some parts of the bedrock are more saturated than other parts, or that the clay content (and thus permeability) of the bedrock differs from location to location. This has implications for the fate of recharge and the distribution of transit times through the bedrock. Although the flysch bedrock is typically assumed to be poorly permeable, fractures in the bedrock might still serve as fast conduits for water at particular locations. This is especially true at locations where tree roots penetrate the bedrock. However, geophysical data is not straightforward to interpret, and uncertainties increase with depth, so more investigations are needed to clarify the role of the bedrock for catchment water storage and release.

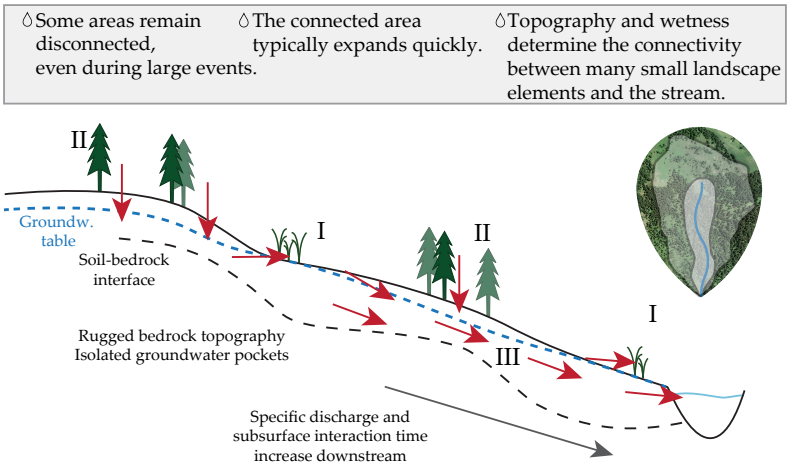

Figure 7.1 – Schematic drawing of a hillslope cross-section (left) with the expected groundwater level at high flow conditions (dashed blue lines) and the soil-bedrock interface (black dashed lines). Red arrows indicate the dominant flow directions and processes (i.e., infiltration, exfiltration or transport), and roman letters (I-III) indicate the different groundwater types. Type I: riparian-like areas with low-permeability soils and shallow groundwater levels that most often contribute to streamflow, Type II: hillslopes with relatively high permeability soils and typically deeper and more variable groundwater levels; these areas become connected to the stream during very wet conditions, and Type III: 'deep' groundwater, present in various soils, especially in the lower parts of the catchment, and connected to the stream depending on the specific location and catchment wetness. And, a catchment (right) with an orthophoto in the background, where the white-shaded close to the stream area represents the area that is permanently connected to the stream, and the white-shaded area farther from the stream area represents the area that is only transiently connected.

Chapter 8

Conclusions

This thesis presents an assessment of the variability in shallow groundwater composition in the pre-alpine Studibach catchment, and how this variability influences stream-based source-area analyses. The main findings can be summarized as follows:

- The spatial variability in shallow groundwater composition in the Studibach is large, and larger than the temporal variability for most elements.
- Four shallow groundwater types could be distinguished based on differences from the catchment average concentrations, but the variability within each type was still large. Three of the four shallow groundwater types could be related to landscape elements.
- Some topographic site attributes were significantly different for the different groundwater types. However, the correlation between solute concentrations and individual site attributes was generally weak. Aquifer chemical composition and local persistence of groundwater levels, controlled by depth-to-bedrock and surface topography, likely have a large effect on the spatial distribution of the shallow groundwater composition.
- Baseflow streamwater chemistry most strongly resembled riparian-like groundwater (Type I) during most campaigns, for all but one subcatchment. However, baseflow was more similar to hillslope

groundwater (Type II) shortly after snowmelt, reflecting differences in hydrologic connectivity.

- The variability in groundwater composition was too large to determine the contributions from the different groundwater types during rainfall events. However, most water was estimated to be groundwater, and soil water contributions were minimal for three of the four analyzed events.
- There was no clear threshold at which streamwater was not a simple mixture of baseflow and precipitation. However, the change in solute concentrations could be used to identify contributions of different groundwater types and soil water to the stormflow mixture.
- The uncertainty in hydrograph separation analyses due to the variable pre-event water composition can be large, and should be acknowledged when interpretating the results of these analyses.
- The uncertainty in hydrograph separation analyses can be reduced by increasing the number of groundwater samples. For the Studibach, the largest improvement occurred when the number of samples was increased from three to six. Taking samples at a few more locations than is typically done can already be helpful in estimating the spatial variability and reducing uncertainties.

Chapter 9

Outlook

This PhD thesis contributes a piece to the puzzle that describes the hydrologic functioning of the Alptal headwater catchments, and yields food for thought for some common approaches in catchment hydrology. From this work, we have learned more about the specific functioning of the Studibach catchment, and tested assumptions and hypotheses regarding the representation of pre-event water in source-area analyses. As is usual after studying a specific subject, one ends up with more questions than at the start. In this chapter, I will suggest two open questions for the field of catchment sciences that I believe are feasible and useful to address.

9.1 Catchment similarity and scaling

Nearby catchments with similar landscape characteristics are often assumed to respond hydrologically similar to rainfall, especially in humid temperate climates (Patil and Stieglitz, 2012). We might ask the question if nearby catchments also have comparable hydrochemical responses. The Alptal is a good location for testing this hypothesis. Different datasets are readily available and encompass measurements at multiple spatial scales, from $\sim$0.1 km^2 (C6, Studibach) up to 0.7 km^2 for the Erlenbach and 83 km^2 for the entire Alptal watershed. Apart from continuous stream stage height measurements at each of these sites, more detailed information

is available for the Erlenbach and Studibach. Three of the baseflow snapshot campaigns were performed simultaneously in the Erlenbach and Studibach, and stormflow sampling also occurred during several rainfall events in both catchments. A comparison of the hydrometric and hydrochemical response of these catchments can therefore yield valuable insights in the similarity or differences in hydrological processes, and might help us understand if and how these processes change from location to location or with increasing catchment size.

Preliminary work on the 2017 data suggests that specific discharge was higher at the Studibach than at the Erlenbach and Alp runoff stations, and that the seasonal fluctuations were more pronounced for the Studibach as well. The correlation between the magnitudes of different events was weakest for the comparison between the Alp and Studibach and strongest for the two Studibach catchments. Streamwater chemistry during the baseflow snapshot campaigns was very similar in the upper parts of the Studibach and Erlenbach, and less similar for the lower part of the Erlenbach, which is at a lower elevation than the Studibach outlet. The response of different solutes during rainfall events was for some tracers very similar in both catchments (e.g., for deuterium; Figure 9.1) but for other tracers differed strongly (e.g., for iron; Figure 9.1).

Hypotheses such as those listed below could be tested with these data.

- Specific discharge decreases linearly with catchment size in the Alptal valley.
- Streamwater chemistry, and thus source areas, in Erlenbach and Studibach are more similar during baseflow conditions than during stormflow conditions.
- Differences in the change in iron concentrations during rain events between Erlenbach and Studibach are related to the percentage of wetland area in each catchment.

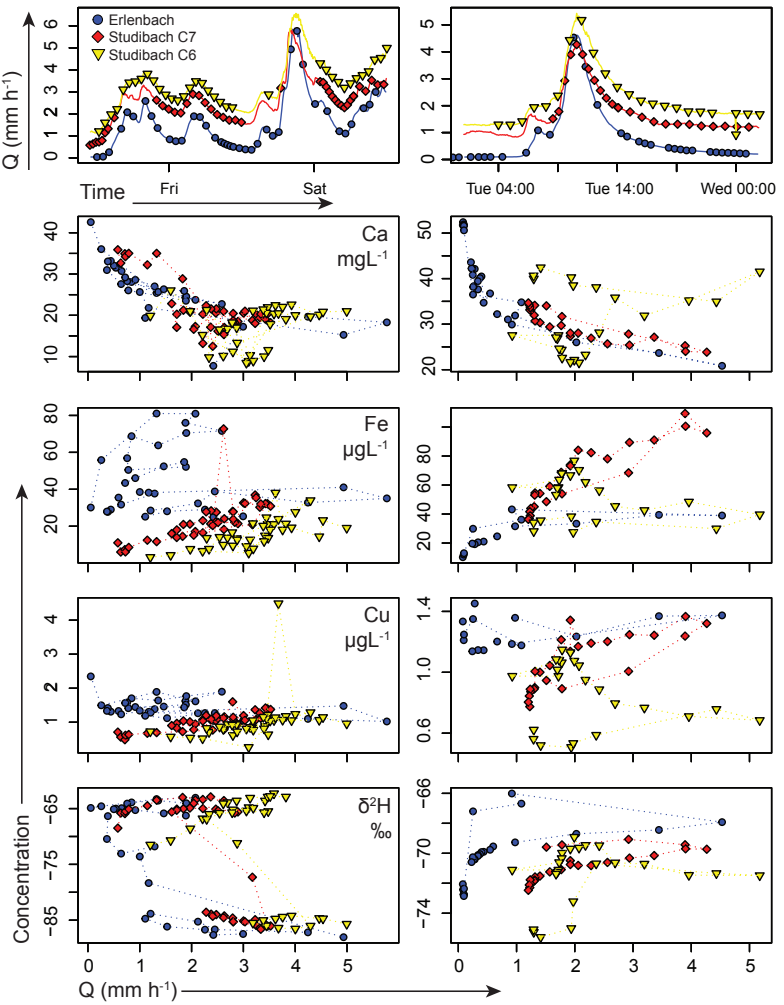

Figure 9.1 – Discharge (Q, mm h^{-1}, line graph) and stormflow sampling times (symbols) for two events at the outlets of the Erlenbach (blue circles), Studibach C7 (red diamonds) and Studibach C6 (yellow reversed triangles) (top row), and concentration discharge plots for calcium, iron, copper and deuterium (remaining four rows).

9.2 Subsurface runoff processes

We know little about what happens below the surface during rain events. Most often we investigate groundwater processes from a stream-based perspective, and use groundwater level observations and changes in stormflow composition to infer groundwater contributions (as was done in this thesis). Switching to a groundwater-perspective could help us to better understand the recharge and drainage processes that drive the changes groundwater levels and how these vary across the landscape. One way to investigate this is by sampling groundwater at different wells during rainfall events and to compare the response at each location.

Preliminary work shows that the groundwater composition changes during rainfall events, but that the changes differ per sampling location (shown for Event III in Figure 9.2). During event III, samples were collected at three Type I wells, one Type II well, one Type III well, and at the outlets of C6 and C7. For two of the Type I wells and the Type III well the composition changed less ($\sim$ 4 ‰ δ^2H and 10 mg L^{-1} calcium) than the composition of the third Type I well and the Type II well ($\sim$ 5 ‰ δ^2H and 30 mg L^{-1} calcium). For comparison, the 4-5 ‰ δ^2H change roughly equals the standard deviation in groundwater during a snapshot campaign (range: 18 ‰ δ^2H), whereas the median range for calcium was 50 (temporal) to 77 (spatial) mg L^{-1} (see paper I).

These changes in groundwater composition suggest fast recharge of the groundwater, and that using established tools such as EMMA or hydrograph separation might be feasible to investigate groundwater recharge sources and dominant subsurface runoff processes. For instance, such data could be used to test the following hypotheses:

- The groundwater composition during rain events is a mixture of soil water and pre-event groundwater.
- The change in concentration of geogenic solutes is negatively related to the relative change in groundwater level (i.e., larger water level changes result in a stronger concentration decrease).

- Concentration changes during rain events at Type III wells are smaller than concentration changes at Type I and Type II wells.

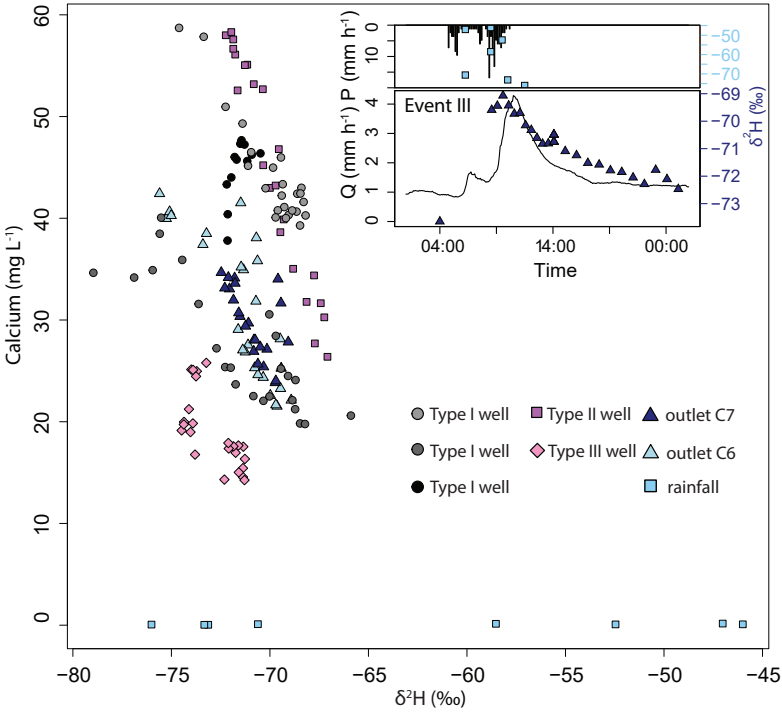

Figure 9.2 – Correlation between the calcium concentration and isotopic composition for groundwater, precipitation and stormflow for a 30-mm rainfall event on 3 October 2017 (Event III). Groundwater Type I (black and grey circles) was sampled at three individual wells, whereas groundwater Type II (purple squares) and III (pink diamonds) were sampled at one well. Insert: The timeseries of precipitation (bar graph) and specific discharge C7 (line graph), and δ^2H of precipitation (light blue squares) and stormflow at C7 (dark blue triangles).

s

Acknowledgements

I could not have imagined how good the decision for this PhD project was going to be for me. I got the opportunity to work on an interesting topic, while being offered the time and resources to develop ideas and to go and apply different methods in the field. Of course, none of this would have been possible without the help of many people, and particularly not without the trust and dedication of Jan and Ilja.

Ilja, I don't think words can describe how helpful and stimulating your supervision was during this PhD. From brainstorming about the sampling strategy, to proof-reading the final work. You always made time to discuss, you guided me back to my research questions numerous times, and you were always positive. But you went much further than this. You checked in on all aspects of my life, and made sure that I was doing well. This made me feel at home in H2K and Switzerland, and helped me through the challenging periods. Thank you so much for all your time, effort and friendship along the way.

Jan, I experienced you to be always enthusiastic. If it was about taking more samples or measurements in the Alptal, about presenting at another meeting, or about doing sports, for any activity I could count on your support. In all aspects where you are supportive, you also take your responsibility for guidance. Hence, you guided me through new experiences like going to scientific conferences, teaching, presenting science to the general public, or dealing with a paper rejection. You have a special way of gathering people, which makes the H2K group an extremely safe and friendly environment - I would argue the perfect environment for doing a PhD. Thank you for offering me this opportunity and supporting me all the way.

Of course there were more people advising and helping me. Manfred, thank you for monitoring my progress, and stimulating me to finalize

specific parts. I sometimes got lost in 'doing another measurement' while actually the story was there, and you helped me to realize this. You were always quick to reply and very helpful, and the time you invest in the Alptal community is invaluable. Markus, you perfectly fulfilled your role as committee member by suggesting out-of-the-box methods during the first part of the PhD, and stimulating me to finish towards the end. Thank you both for your time and commitment.

H2K was really like a family for me during the PhD, and made the four years fly by with everyday discussions over coffee and days out in the mountains. Thank you Barbara, for your friendship and teaching me how to climb a multipitch, Simon, for helping me to grow spiritually and going on mountaineering adventures, Jana, for your positive and plant-loving attitude, Fabian, for being available to listen to me at all times, and for making L90 the best office ever, Kirsti, for being honest and for being my view in the office for four years, Manuela, for showing me how to work efficiently, Sandra, for teaching me Swiss-German, Ania, for the funny coffee-chats, Ling for the kind and caring words, Maria, for always making jokes, Marc, for giving great hiking tips and fixing my keyboards, Rick, for highlighting the fun aspect in many situations, Dani, for the interesting discussions over lunch, Daphné, for bringing joy and inspiration, Florian, for starting breaks at the exact time that we agreed on, Ben, for showing me the magic of the Alptal, Michi, for discussing about and advising me on the Alptal groundwaters, and Isabelle and Frank, for being extremely kind and attentive. It was also great to spend time with the extended GIUZ-family, especially with Alina and Ale, but of course many other kind individuals were also there along the way.

Working in the Alptal is a privilege. The field site is beautiful, and the community of people working in the Alptal is equally special. Jana, you helped me tremendously with discussions about the conceptual model of the Alptal, and by spending your time wrestling through the messy data. Julia, your positive and efficient work attitude helped moving the event-based analyses and interpretations forward. James, you know how to ask the one question that gets to the core of the problem, and that challenged and motivated me. Bjorn, thank you for your time in the lab and for explaining me how to do measurements. Kari, thanks for helping me to unscrew the hand-auger whenever it was stuck.

I also found that being out in the field with other Alptal-lovers such as Andrea, Toby, Nadja and Anna was great fun, and made the days in the Alptal fly by. Special thanks to Adrian, for organizing the geophysical measurements, but also to all the persons that assisted with the field measurements over the years: performing geophysical measurements, augering, sampling of water, soil and rocks, installing the weather-station, installing suction lysimeters, reading out loggers, finding loggers in the

snow, carrying ISCO's or other heavy items up the steep slopes, and all of that under the assumption that there would be a nice view. Thank you all for your time and effort!

After coming back from the Alptal there was still laboratory work to do. I would like to especially thank Barbara for handling many, many isotope samples in the Freiburg lab, Christophe, for inviting me to visit LIST and helping me to do the leaching experiments, and the GIUZ-lab-crew that accomodated me during the many hours of filtering water samples, and fiddling with ISCO-batteries.

Many hours I spent also with my rugby family. They maybe did not always understand why 'I was catching rain', but they were always there for me. Nothing tops Tatjana's morning coffee's or Anna's cinnamon rolls in times of despair. The same goes for my parents and brother, who have eagerly listened to stories about groundwater and asked me about updates on the papers, and my friends back home in The Netherlands. You were always with me in my mind, and your support enabled me to grow beyond what I thought would be possible. And finally, when coming home at the end of the day, there was always the WG-community that I could count on. Living with Ivan, Bolesch, Marion, Alyosha, Justas, Haensel, Flavia, Stéphane and Andi changed my life for the better, and was all that I could wish for when arriving in a new city at the start of the PhD. You made me feel at home, supported me in tough periods, and were there to make it all fun. And what grew from living together was even more special. Ivan, you have become one of the most important persons in my life. You stimulate me to be a better individual, and you help me to achieve more in every possible way. Your support and love optimized the circumstances for my PhD and I'm sure it will be the same for any project that follows.

Bibliography

Abbott, B. W., G. Gruau, J. P. Zarnetske, F. Moatar, L. Barbe, Z. Thomas, O. Fovet, T. Kolbe, S. Gu, A. C. Pierson-Wickmann, P. Davy, and G. Pinay
2018. Unexpected spatial stability of water chemistry in headwater stream networks. *Ecology Letters*, 21(2):296–308.

Aubert, A. H., C. Gascuel-Odoux, G. Gruau, N. Akkal, M. Faucheux, Y. Fauvel, C. Grimaldi, Y. Hamon, A. Jaffrézic, M. Lecoz-Boutnik, J. Molénat, P. Petitjean, L. Ruiz, and P. Merot
2013. Solute transport dynamics in small, shallow groundwater-dominated agricultural catchments: insights from a high-frequency, multisolute 10 yr-long monitoring study. *Hydrology and Earth System Sciences*, 17(4):1379–1391.

Bachmair, S. and M. Weiler
2012. Hillslope characteristics as controls of subsurface flow variability. *Hydrology and Earth System Sciences*, 16(10):3699–3715.

Barthold, F. K., C. Tyralla, K. Schneider, K. B. Vaché, H. G. Frede, and L. Breuer
2011. How many tracers do we need for end member mixing analysis (EMMA)? A sensitivity analysis. *Water Resources Research*, 47(8):1–14.

Beven, K. J. and M. J. Kirkby
1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1):43–69.

Birkel, C. and C. Soulsby
2015. Advancing tracer-aided rainfall-runoff modelling: A review of progress, problems and unrealised potential. *Hydrological Processes*, 29(25):5227–5240.

Blume, T. and H. I. van Meerveld
2015. From hillslope to stream: methods to investigate subsurface connectivity. *Wiley Interdisciplinary Reviews: Water*, 2(3):177–198.

Bonell, M.
1998. Selected challenges in runoff generation research in forests from the hillslope to headwater drainage basin scale. *Journal of the American Water Resources Association*, 34(4):765–785.

Bracken, L. J. and J. Croke
2007. The concept of hydrological connectivity and its contribution to understanding runoff-dominated geomorphic systems. *Hydrological Processes*, 21:2267–2274.

Brammer, D. and J. McDonnell
1996. An evolving perceptual model of hillslope flow at the maimai catchment. *Advances in Hillslope Processes*, 1:35–60.

Brown, V. A., J. J. McDonnell, D. A. Burns, and C. Kendall
1999. The role of event water, a rapid shallow flow component, and catchment size in summer stormflow. *Journal of Hydrology*, 217(3-4):171–190.

Brunke, M. and T. Gonser
1997. The ecological significance of exchange processes between rivers and groundwater. *Freshwater biology*, 37(February 1997):1–33.

Burns, D. A., R. P. Hooper, J. J. McDonnell, J. E. Freer, C. Kendall, and K. Beven
1998. Base cation concentrations in subsurface flow from a forested hillslope: The role of flushing frequency. *Water Resources Research*, 34(12):3535–3544.

Burns, D. A., L. N. Plummer, J. J. McDonnell, E. Busenberg, G. C. Casile, C. Kendall, R. P. Hooper, J. E. Freer, N. E. Peters, K. Beven, and P. Schlosser
2003. The Geochemical Evolution of Riparian Ground Water in a Forested Piedmont Catchment. *Groundwater*, 41(7):913–925.

Buttle, J.
1994. Isotope hydrograph separations and rapid delivery of pre-event water from drainage basins. *Progress in Physical Geography: Earth and Environment*, 18:16–41.

Carey, S. K. and W. L. Quinton
2005. Evaluating runoff generation during summer using hydrometric, stable isotope and hydrochemical methods in a discontinuous permafrost alpine catchment. *Hydrological Processes*, 19(1):95–114.

Cayuela, C., J. Latron, J. Geris, and P. Llorens
2019. Spatio-temporal variability of the isotopic input signal in a partly forested catchment: Implications for hydrograph separation. *Hydrological Processes*, 33(1):36–46.

Christophersen, N. and R. P. Hooper
1992. Multivariate Analysis of Stream Water Chemical Data' The Use of Principal Components Analysis for the End-Member Mixing Problem. *Water Resources Research*, 28(1):99–107.

Detty, J. M. and K. J. McGuire
2010. Topographic controls on shallow groundwater dynamics: Implications of hydrologic connectivity between hillslopes and riparian zones in a till mantled catchment. *Hydrological Processes*, 24(16):2222–2236.

Fischer, B., F. Aemisegger, P. Graf, H. Sodemann, and J. Seibert
2019. Assessing the Sampling Quality of a Low-Tech Low-Budget Volume-Based Rainfall Sampler for Stable Isotope Analysis. *Frontiers in Earth Science*, 7:1–8.

Fischer, B. M., H. J. van Meerveld, and J. Seibert
2017. Spatial variability in the isotopic composition of rainfall in a small headwater catchment and its effect on hydrograph separation. *Journal of Hydrology*, 547:755–769.

Fischer, B. M. C., M. Rinderer, P. Schneider, T. Ewen, and J. Seibert
2015. Contributing sources to baseflow in pre-alpine headwaters using spatial snapshot sampling. *Hydrological Processes*, 29(26):5321–5336.

Fischer, B. M. C., M. Stahli, and J. Seibert
2016. Pre-event water contributions to runoff events of different magnitude in pre-alpine headwaters. *Hydrology Research*, 48:28–47.

Genereux, D.
1998. Quantifying uncertainty in tracer-based hydrograph separations. *Water Resources Research*, 34(4):915–919.

Hagedorn, F., P. Schleppi, P. Waldner, and H. Flühler
2000. Export of dissolved organic carbon and nitrogen from Gleysol dominated catchments–the significance of water flow paths. *Biogeochemistry*, 50:137–161.

Harter, R.
1983. Effect of soil pH on adsorption of lead, copper, zinc, and nickel. *Soil Science Society of America Journal*, (47):47–51.

Haught, D. R. W. and H. J. van Meerveld
2011. Spatial variation in transient water table responses: Differences between an upper and lower hillslope zone. *Hydrological Processes*, 25(25):3866–3877.

Hegg, C., B. W. McArdell, and A. Badoux
2006. One hundred years of mountain hydrology in Switzerland by the WSL. *Hydrological Processes*, 20(2):371–376.

Hissler, C., P. Stille, J. Juilleret, J. François, T. Perrone, and G. Morvan
2015. Elucidating the formation of terra fuscas using Sr – Nd – Pb isotopes and rare earth elements. *Applied Geochemistry*, 54:85–99.

Hooper, R. P., N. Christophersen, and N. E. Peters
1990. Modelling streamwater chemistry as a mixture of soilwater end-members - An application to the Panola Mountain catchment, Georgia, U.S.A. *Journal of Hydrology*, 116(1-4):321–343.

Hugenschmidt, C., J. Ingwersen, W. Sangchan, Y. Sukvanachaikul, A. Duffner, S. Uhlenbrook, and T. Streck
2014. A three-component hydrograph separation based on geochemical tracers in a tropical mountainous headwater catchment in northern Thailand. *Hydrology and Earth System Sciences*, 18(2):525–537.

Inamdar, S. P. and M. J. Mitchell
2006. Hydrologic and topographic controls on storm-event exports of dissolved organic carbon (BOC) and nitrate across catchment scales. *Water Resources Research*, 42(3):1–16.

Jencso, K. G. and B. L. McGlynn
2011. Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation. *Water Resources Research*, 47(11):1–16.

Jencso, K. G., B. L. McGlynn, M. N. Gooseff, K. E. Bencala, and S. M. Wondzell
2010. Hillslope hydrologic connectivity controls riparian groundwater turnover: Implications of catchment structure for riparian buffering and stream water sources. *Water Resources Research*, 46(10):1–18.

Jencso, K. G., B. L. McGlynn, M. N. Gooseff, K. E. Bencala, S. M. Wondzell, K. E. Bencala, and L. A. Marshall
2009. Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale. *Water Resources Research*, 45(4):1–16.

Kaushal, S. S., A. J. Gold, S. Bernal, T. A. Newcomer Johnson, K. Addy, A. Burgin, D. A. Burns, A. A. Coble, E. Hood, Y. Lu, P. Mayer, E. C. Minor, A. W. Schroth, P. Vidon, H. Wilson, M. A. Xenopoulos, T. Doody, J. G. Galella, P. Goodling, K. Haviland, S. Haq, B. Wessel, K. L. Wood, N. Jaworski, and K. T. Belt
2018. Watershed 'chemical cocktails': forming novel elemental combinations in Anthropocene fresh waters. *Biogeochemistry*, 141(3):281–305.

Kendall, C. and J. J. Mcdonnell
1998. *Isotope Tracers in Catchment Hydrology.* Amsterdam: Elsevier Science Publishers.

Kendall, C., J. J. McDonnell, and W. Gu
2001. A look inside 'black box' hydrograph separation models: A study at the hydrohill catchment. *Hydrological Processes*, 15(10):1877–1902.

Kennedy, V. C., G. W. Zellweger, and R. J. Avanzino
1979. Variation of rain chemistry during storms at two sites in northern California. *Water Resources Research*, 15(3):687–702.

Klaus, J., K. K. Chun, K. J. McGuire, and J. J. Mcdonnell
2015a. Temporal dynamics of catchment transit times from stable isotope data. *Water Resources Research*, 51:4208–4223.

Klaus, J. and J. J. McDonnell
2013. Hydrograph separation using stable isotopes: Review and evaluation. *Journal of Hydrology*, 505:47–64.

Klaus, J., J. J. McDonnell, C. R. Jackson, E. Du, and N. A. Griffiths
2015b. Where does streamwater come from in low-relief forested watersheds? A dual-isotope approach. *Hydrology and Earth System Sciences*, 19(1):125–135.

Knapp, J. L. A., J. von Freyberg, B. Studer, L. Kiewiet, and J. W. Kirchner
2020. Concentration-discharge relationships vary among hydrological events, reflecting differences in event characteristics. *Hydrology and Earth System Sciences Discussions*, 2020:1–27.

Laudon, H. and O. Slaymaker
1997. Hydrograph separation using stable isotopes, silica and electrical conductivity: An alpine example. *Journal of Hydrology*, 201(1-4):82–101.

Legout, C., J. Molenat, L. Aquilina, C. Gascuel-Odoux, M. Faucheux, Y. Fauvel, and T. Bariac"
"2007". "solute transfer in the unsaturated zone-groundwater continuum of a headwater catchment". *"Journal of Hydrology"*, "332"("3"):"427 – 441".

Lindsay, J. B.
2016. Whitebox GAT: A case study in geomorphometric analysis. *Computers and Geosciences*, 95:75–84.

Loheide, S. P. and S. M. Gorelick
2006. Quantifying stream-aquifer interactions through the analysis of remotely sensed thermographic profiles and in situ temperature histories. *Environmental Science and Technology*, 40(10):3336–3341.

Lyon, S. W., S. L. E. Desilets, and P. A. Troch
2009. A tale of two isotopes: differences in hydrograph separation for a runoff event when using d versus 18o. *Hydrological Processes*, 23(14):2095–2101.

Martin, C., L. Aquilina, C. Gascuel-Odoux, J. Molénat, M. Faucheux, and L. Ruiz
2004. Seasonal and interannual variations of nitrate and chloride in stream waters related to spatial and temporal patterns of groundwater concentrations in agricultural catchments. *Hydrological Processes*, 18(7):1237–1254.

McDonnell, J., M. Stewart, and I. Owens
1991. Effect of Catchment-Scale Subsurface Mixing on Stream Isotopic Response. *Water Resources Research*, 27(12):3065–3073.

McGlynn, B. L. and J. J. McDonnell
2003. Quantifying the relative contributions of riparian and hillslope zones to catchment runoff. *Water Resources Research*, 39(11):1310.

McGlynn, B. L. and J. Seibert
2003. Distributed assessment of contributing area and riparian buffering along stream networks. *Water Resources Research*, 39(4):1–7.

McKenzie, R. M.
1980. The adsorption of lead and other heavy metals on oxides of manganese and iron. *Australian Journal of Soil Research*, 18(1):61–73.

Megonigal, J. P., W. H. Patrick, and S. P. Faulkner
1993. Wetland Identification in Seasonally Flooded Forest Soils: Soil Morphology and Redox Dynamics. *Soil Science Society of America Journal*, 57(1):140.

Mohn, J., A. Schürmann, F. Hagedorn, P. Schleppi, and R. Bachofen
2000. Increased rates of denitrification in nitrogen-treated forest soils. *Forest Ecology and Management*, 137(1-3):113–119.

Molnar, P., A. L. Densmore, B. W. McArdell, J. M. Turowski, and P. Burlando
2010. Analysis of changes in the step-pool morphology and channel profile of a steep mountain stream following a large flood. *Geomorphology*, 124(1-2):85–94.

Patil, S. and M. Stieglitz
2012. Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment. *Hydrology and Earth System Sciences*, 16(2):551–562.

Pellerin, B. A., W. M. Wollheim, X. Feng, and J. V. Charles
2008. The application of electrical conductivity as a tracer for hydrograph separation in urban catchments. *Hydrological Processes*, 22:1810–1818.

Penna, D., N. Mantese, L. Hopp, G. Dalla Fontana, and M. Borga
2015. Spatio-temporal variability of piezometric response on two steep alpine hillslopes. *Hydrological Processes*, 29(2):198–211.

Penna, D. and H. van Meerveld
2019. Spatial variability in the isotopic composition of water in small catchments and its effect on hydrograph separation. *WIREs Water*, Pp. 1–33.

Penna, D., H. J. van Meerveld, G. Zuecco, G. Dalla Fontana, and M. Borga
2016. Hydrological response of an Alpine catchment to rainfall and snowmelt events. *Journal of Hydrology*, 537:382–397.

Penna, D., G. Zuecco, S. Crema, S. Trevisani, M. Cavalli, L. Pianezzola, L. Marchi, and M. Borga
2017. Response time and water origin in a steep nested catchment in the Italian Dolomites. *Hydrological Processes*, 31(4):768–782.

Rinderer, M., H. van Meerveld, and J. Seibert
2014. Topographic controls ond shallow groundwater levels in a steep, prealpine catchment. *Water Resources Research*, (50):6067–6080.

Rinderer, M., H. J. van Meerveld, and B. L. McGlynn
2019. From points to patterns: Using groundwater time series clustering to investigate subsurface hydrological connectivity and runoff source area dynamics. *Water Resources Research*, 55(7):5784–5806.

Rinderer, M., I. van Meerveld, M. Stähli, and J. Seibert
2015. Is groundwater response timing in a pre-alpine catchment controlled more by topography or by rainfall? *Hydrological Processes*, 30(7):1036–1051.

Rodhe, A.
1987. *The Origin of Stream Traced by Oxygen 18*. Report series a 41, Uppsala University.

Schleppi, P., N. Muller, H. Feyen, A. Papritz, J. B. Bucher, and H. Flühler
1998. Nitrogen budgets of two small experimental forested catchments at Alptal , Switzerland. *Forest Ecology and Management*, 101(1-3):177–185.

Schneider, P., S. Pool, L. Strouhal, and J. Seibert
2014. True colors-experimental identification of hydrological processes at a hillslope prone to slide. *Hydrology and Earth System Sciences*, 18(2):875–892.

Seibert, J., K. Bishop, L. Nyberg, and A. Rodhe
2011. Water storage in a till catchment. I: Distributed modelling and relationship to runoff. *Hydrological Processes*, 25(25):3937–3949.

Seibert, J., K. Bishop, A. Rodhe, and J. J. McDonnell
2003. Groundwater dynamics along a hillslope: A test of the steady state hypothesis. *Water Resources Research*, 39(1):1–9.

Seibert, J., T. Grabs, S. Köhler, H. Laudon, M. Winterdahl, and K. Bishop
2009. Linking soil-and stream-water chemistry based on a Riparian Flow-Concentration Integration Model. *Hydrol. Earth Syst. Sci*, 13:2287–2297.

Shotyk, W.
1996. Natural and anthropogenic enrichments of As, Cu, Pb, Sb, and Zn in ombrotrophic versus minerotrophic peat bog profiles, Jura Mountains, Switzerland. *Water, Air, and Soil Pollution*, 90(3-4):375–405.

Shotyk, W., P. Blaser, A. Grünig, and A. K. Cheburkin
2000. A new approach for quantifying cumulative, anthropogenic, atmospheric lead deposition using peat cores from bogs: Pb in eight Swiss peat bog profiles. *Science of the Total Environment*, 249(1-3):281–295.

Sidle, R. C., Y. Tsuboyama, S. Noguchi, I. Hosoda, M. Fujieda, and T. Shimizu
2000. Stormflow generation in steep forested headwaters: a linked hydrogeomorphic paradigm. *Hydrological Processes*, 14:369–385.

Silliman, S. E., B. Berkowitz, J. Simunek, and M. T. van Genuchten
2002. Fluid flow and solute migration within the capillary fringe. *Groundwater*, 40(1):76–84.

Sklash, M., R. Farvolden, and P. Fritz
1976. A conceptual model of watershed response to rainfall, developed through the use of oxygen-18 as a natural tracer. *Canadian Journal of Earth Sciences*, 13(2):271–283.

Soulsby, C., M. Chen, R. C. Ferrier, R. C. Helliwell, A. Jenkins, and R. Harriman
1998. Hydrogeochemistry of shallow groundwater in an upland Scottish catchment. *Hydrological Processes*, 12(7):1111–1127.

Soulsby, C., D. Tetzlaff, P. Rodgers, S. Dunn, and S. Waldron
2006. Runoff processes, stream water residence times and controlling landscape characteristics in a mesoscale catchment: An initial evaluation. *Journal of Hydrology*, 325(1-4):197–221.

Stähli, M.
2018. Longterm hydrological observatory Alptal (central Switzerland).

Stähli, M. and D. Gustafsson
2006. Long-term investigations of the snow cover in a subalpine semi-forested catchment. *Hydrological Processes*, 20(2):411–428.

Stieglitz, M., J. Shaman, J. McNamara, V. Engel, J. Shanley, and G. W. Kling
2003. An approach to understanding hydrologic connectivity on the hillslope and the implications for nutrient transport. *Global Biogeochemical Cycles*, 17(4):n/a–n/a.

Tague, C., G. Grant, M. Farrell, J. Choate, and A. Jefferson
2008. Deep groundwater mediates streamflow response to climate warming in the Oregon Cascades. *Climatic Change*, 86(1-2):189–210.

Tetzlaff, D. and C. Soulsby
2008. Sources of baseflow in larger catchments - Using tracers to develop a holistic understanding of runoff generation. *Journal of Hydrology*, 359(3-4):287–302.

Tromp-van Meerveld, H. J. and J. J. McDonnell
2006. Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, 42(2):1–11.

van Meerveld, H., B. Fischer, M. Rinderer, M. Stähli, and J. Seibert
2018. Runoff generation in a pre-alpine catchment: A discussion between a tracer and a shallow groundwater hydrologist. *Cuadernos de Investigación Geográfica*, 44:429–452.

van Meerveld, H. J., J. Seibert, and N. E. Peters
2015. Hillslope-riparian-stream connectivity and flow directions at the Panola Mountain Research Watershed. *Hydrological Processes*, 29(16):3556–3574.

van Meerveld, H. J. I., J. W. Kirchner, M. J. P. Vis, R. S. Assendelft, and J. Seibert
2019. Expansion and contraction of the flowing stream network alter hillslope flowpath lengths and the shape of the travel time distribution. *Hydrology and Earth System Sciences*, 23(11):4825–4834.

von Freyberg, J., D. Radny, H. E. Gall, and M. Schirmer
2014. Implications of hydrologic connectivity between hillslopes and riparian zones on streamflow composition. *Journal of Contaminant Hydrology*, 169:62–74.

von Freyberg, J., D. Radny, H. E. Gall, and M. Schirmer
2014. Implications of hydrologic connectivity between hillslopes and riparian zones on streamflow composition. *Journal of Contaminant Hydrology*, 169:62–74.

von Freyberg, J., B. Studer, and J. W. Kirchner
2017. A lab in the field: high-frequency analysis of water quality and stable isotopes in stream water and precipitation. *Hydrology and Earth System Sciences*, 21(3):1721–1739.

von Freyberg, J., B. Studer, M. Rinderer, and J. W. Kirchner
2018. Studying catchment storm response using event- and pre-event-water volumes as fractions of precipitation rather than discharge. *Hydrology and Earth System Sciences*, 22(11):5847–5865.

Walker, J. F., R. J. Hunt, T. D. Bullen, D. P. Krabbenhoft, and C. Kendall
2003. Variability of Isotope and Major Ion Chemistry in the Allequash Basin, Wisconsin. *GroundWater*, 41(7):883–894.

Welsch, D. L., C. N. Kroll, J. J. McDonnell, and D. A. Burns
2001. Topographic controls on the chemistry of subsurface stormflow. *Hydrological Processes*, 15(10):1925–1938.

Appendix

A: Paper I

Received: 4 August 2018 | Accepted: 20 May 2019
DOI: 10.1002/hyp.13517

SI JAKE PETERS

WILEY

Spatiotemporal variability in hydrochemistry of shallow groundwater in a small pre-alpine catchment: The importance of landscape elements

Leonie Kiewiet[1] | Jana von Freyberg[2,3] | H.J. (Ilja) van Meerveld[1]

[1]Department of Geography, University of Zürich, Zürich, Switzerland

[2]Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland

[3]Mountain Hydrology and Mass Movements, Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

Correspondence
Leonie Kiewiet, Hydrology and Climate Unit, Department of Geography, University of Zürich, Winterthurerstr. 190, 8057CH Zürich, Switzerland.
Email: leonie.kiewiet@geo.uzh.ch

Abstract

Topography and landscape characteristics affect the storage and release of water and, thus, groundwater dynamics and chemistry. Quantification of catchment scale variability in groundwater chemistry and groundwater dynamics may therefore help to delineate different groundwater types and improve our understanding of which parts of the catchment contribute to streamflow. We sampled shallow groundwater from 34 to 47 wells and streamflow at seven locations in a 20-ha steep mountainous catchment in the Swiss pre-Alps, during nine baseflow snapshot campaigns. The spatial variability in electrical conductivity, stable water isotopic composition, and major and trace ion concentrations was large and for almost all parameters larger than the temporal variability. Concentrations of copper, zinc, and lead were highest at sites that were relatively dry, whereas concentrations of manganese and iron were highest at sites that had persistent shallow groundwater levels. The major cation and anion concentrations were only weakly correlated to individual topographic or hydrodynamic characteristics. However, we could distinguish four shallow groundwater types based on differences from the catchment average concentrations: riparian zone-like groundwater, hillslopes and areas with small upslope contributing areas, deeper groundwater, and sites characterized by high magnesium and sulfate concentrations that likely reflect different bedrock material. Baseflow was not an equal mixture of the different groundwater types. For the majority of the campaigns, baseflow chemistry most strongly resembled riparian-like groundwater for all but one subcatchment. However, the similarity to the hillslope-type groundwater was larger shortly after snowmelt, reflecting differences in hydrologic connectivity. We expect that similar groundwater types can be found in other catchments with steep hillslopes and wet areas with shallow groundwater levels and recommend sampling of groundwater from all landscape elements to understand groundwater chemistry and groundwater contributions to streamflow.

KEYWORDS

Alptal, baseflow, clustering, flushing frequency, isotopes, shallow groundwater, spatial variability, topography

 © 2019 John Wiley & Sons, Ltd.

1 | INTRODUCTION

Streamflow in undisturbed catchments in humid temperate climates consists predominantly of groundwater, even during rainfall events (Buttle, 1994; Klaus & McDonnell, 2013; Laudon & Slaymaker, 1997). Groundwater sustains baseflow (Kendall & Mcdonnell, 1998; Soulsby et al., 1998; Tetzlaff et al., 2007) and affects streamwater quality because it provides an influx of chemical compounds that were acquired along the subsurface flow pathway (Soulsby, Tetzlaff, Rodgers, Dunn, & Waldron, 2006). Groundwater contributions to streamflow also affect stream temperature (Loheide & Gorelick, 2006). Despite the importance of groundwater for streamflow, only a few studies have focused on the spatial variability in its chemical composition in small headwater catchments by directly sampling multiple wells or piezometers (Welsch, Kroll, McDonnell, & Burns, 2001), trench outflow (Burns et al., 1998), or a combination of both (Kendall, McDonnell, & Gu, 2001).

The chemical composition of groundwater at a certain location is controlled by the soil and bedrock material, the contact time with the soil or bedrock, biological processes, and/or mixing with different water sources from the surface and subsurface. Consequently, groundwater chemistry is expected to vary throughout the catchment. Hillslope and riparian groundwater chemistry tend to differ due to differences in biogeochemical processes (e.g., Cirmo & McDonnell, 1997) and the accumulation of organic material in the riparian zone (Lidman, Boily, Laudon, & Köhler, 2017; McGlynn, McDonnell, Stewart, & Seibert, 2003). However, even riparian zone groundwater chemistry can vary significantly over very short distances. For example, Penna, van Meerveld, Zuecco, Dalla Fontana, and Borga (2016) showed for the Bridge Creek catchment in the Italian Dolomites that the isotopic composition and electrical conductivity (EC) of groundwater in two wells, which were only a few meters apart, were very different.

Several studies hypothesized that a relationship between solute concentrations or isotopic composition in groundwater and subsurface flow regimes exists, but none of them tested this directly by comparing concentrations to groundwater level dynamics. For instance, Welsch et al. (2001) found a strong positive correlation between nitrate concentrations in streamwater and groundwater and the topographic wetness index (TWI; Beven & Kirkby, 1979) for six rainfall events. The nitrate concentrations also varied with soil temperature and antecedent precipitation, and thus, they hypothesized that nitrate-rich groundwater is flushed from the soils into the streams, with the soil wetness being a controlling factor on nitrate concentrations in groundwater. Burns et al. (1998) found an inverse relation between the accumulated bedrock area and base cation concentrations in subsurface flow for the trenched hillslope at the Panola Mountain Research Watershed and also related this to flushing frequency: Subsurface flow occurred more often in areas with a high bedrock accumulated area, which caused these areas to be flushed more frequently, resulting in lower base cation concentrations in subsurface flow. Kendall et al. (2001) reported a widespread variability in the isotopic composition of all subsurface water components for the artificial hydrohill catchment and attributed this variability to differences in flow conditions through macropore flow or matrix flow. Weiler and McDonnell (2006) tested a similar hypothesis through virtual experiments and showed that even if groundwater nutrient concentrations were uniform across a 100-m hillslope at the beginning of the simulations, simulated concentrations were highly variable after five events. They also showed that most of the nutrients in streamwater originated from the bottom of the hillslope and that antecedent wetness conditions strongly controlled nutrient fluxes to the stream due to the higher hydraulic conductivity in the upper part of the soil profile. Lastly, Walker, Hunt, Bullen, Krabbenhoft, and Kendall (2003) measured the chemical variability of groundwater below the streambed and in three hillslope piezometers and found a large spatial variability in major element concentrations, $\delta^{18}O$ and strontium isotopes. Simulated particle tracking of Walker et al.'s data (Pint, Hunt, & Anderson, 2003) suggested that this high variability in groundwater chemistry can be explained by converging flow lines, resulting in outflow of very different groundwater sources in close proximity.

Although it has been recognized that groundwater chemistry is variable throughout catchments and that groundwater storages are rarely well mixed (e.g., Kirchner, 2003; Kirchner & Neal, 2013), small headwater catchment studies commonly assume that groundwater has a more or less constant chemical composition. Groundwater is often sampled at only a few locations or its chemical composition is inferred from baseflow (e.g., Asano, Uchida, Mimasu, & Ohte, 2009; Peralta-Tapia et al., 2015; Stutter, Deeks, Low, & Billett, 2006), which is assumed to reflect a mixture or average of the groundwater in the catchment. For instance, Hugenschmidt et al. (2014) used stream baseflow during rainless periods to define the groundwater signature for three-component hydrograph separation for a 7-km^2 catchment in Thailand. Similarly, in the 2.8-km^2 Shelter Creek catchment, Brown, McDonnell, Burns, and Kendall (1999) used baseflow samples to characterize the pre-event water end-member for a two-component hydrograph separation and showed that these were chemically similar to seep and groundwater samples collected during baseflow, and the transient groundwater near the C-horizon at times of higher baseflow. However, other studies have shown clear differences between the isotopic composition of near-stream groundwater and baseflow (Klaus & McDonnell, 2013). For instance, Burns and McDonnell (1998) compared the seasonal changes in $\delta^{18}O$ in streamwater, soil water, and groundwater in the Panola Mountain Research Watershed and found that in some cases, the seasonal response of the groundwater lagged behind the response of streamwater. Walker et al. (2003) found that near-stream groundwater, sampled in piezometers at two cross-sections (located 0.9, 1.4, and 10 m away from the stream centre for each cross-section) and 1 m below the streambed, differed markedly from the streamwater composition.

Which groundwater source areas contribute to streamflow depends on the hydrological connectivity of the catchment, which

changes depending on the hydrological conditions (Detty & McGuire, 2010; Jencso et al., 2009; von Freyberg, Radny, Gall, & Schirmer, 2014; Zuecco, Rinderer, Penna, Borga, & van Meerveld, 2019). Not all areas of the catchment are connected to the stream during baseflow conditions (e.g., Bracken & Croke, 2007; Jencso & McGlynn, 2011; Ocampo, Sivapalan, & Oldham, 2006). Consequently, the chemical signature of baseflow might resemble only the signature of the riparian and near-stream groundwater sources. When initially disconnected parts of the landscape become hydrologically connected during rainfall or snowmelt events, groundwater with a different chemical signature may start to contribute to streamflow. Kirchner (2003) described this as "old" water storages that contribute in different proportions at high and low flows. Hence, understanding spatial variability in shallow groundwater chemistry is essential to determine source water inputs to streamflow with tracer-based approaches.

Given the limits in our understanding of the spatial variability in groundwater chemistry and the existence of different groundwater stores across a catchment, the specific research questions for this study were as follows:

1. How much does shallow groundwater chemistry vary spatially?
2. Can the spatial variability in shallow groundwater chemistry be explained by topography, shallow groundwater level dynamics, or landscape position?
3. Does streamflow reflect a particular shallow groundwater type, or is it a mixture of all shallow groundwater? And how does this vary seasonally?

1.1 | Study site description

This study took place in the 20-ha Studibach catchment, a headwater catchment of the Zwäckentobel, located in the Alptal, about 40 km southeast of Zürich, Switzerland (N47.038, E8.723). The elevation of the pre-alpine catchment ranges from 1,270 to 1,650 m above sea level (Figure 1). Mean annual precipitation is 2,300 mm/year (Stähli & Gustafsson, 2006). Precipitation is relatively evenly distributed throughout the year (Feyen, Wunderli, Wydler, & Papritz, 1999). During the snow-free season (June to October), it rains on average every second day (van Meerveld, Fischer, Rinderer,

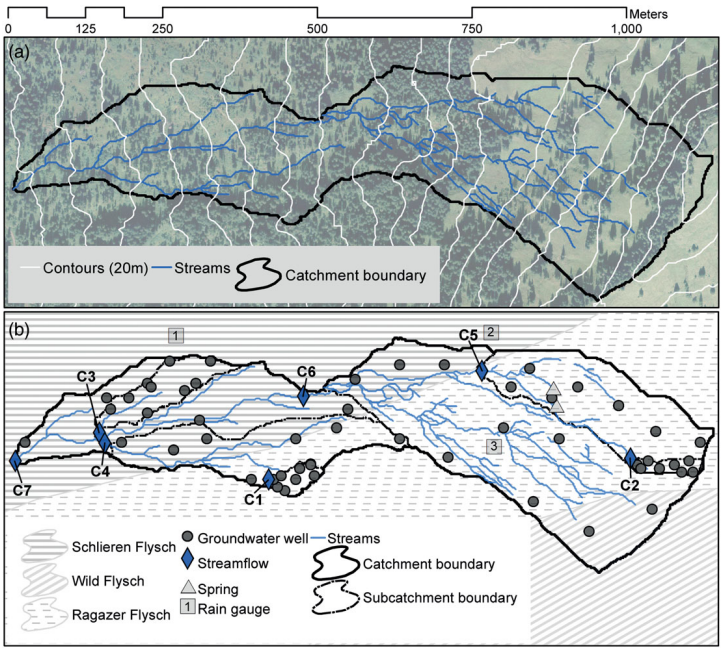

FIGURE 1 Maps of the research area with (a) an aerial photograph, contour lines (20-m interval), streams (blue), and catchment boundary (black), and (b) the catchment (black) and subcatchments (dark grey) boundaries, streams (blue), and streamflow gauges (C1–C7; blue diamonds), springs (light grey triangles), groundwater wells (dark grey dots), and rain gauges (light grey squares), as well as the three main types of Flysch (shading). Source geologic map (GK500) and aerial photograph (aerial image no. 20000090712703): Swiss Federal Office of Topography. Reproduced with permission of Swisstopo (BA18702)

Stähli, & Seibert, 2018). Streamflow and groundwater levels respond quickly to rainfall (Fischer, Rinderer, Schneider, Ewen, & Seibert, 2015; Rinderer, van Meerveld, Stähli, & Seibert, 2015). Median groundwater levels (Rinderer, van Meerveld, & Seibert, 2014) and groundwater response times (Rinderer et al., 2015) are correlated with surface topography and are best described by slope, curvature, and the TWI. The correlations between groundwater levels and the TWI are strongest after peak flow and decrease at the beginning of rainfall events, indicating spatial differences in the groundwater response and connectivity (Rinderer et al., 2014).

The Studibach is underlain by three different types of Flysch (Schlieren Flysch, Ragazer Flysch, and Wild Flysch), which are all of sedimentary and calcareous origin (Cretaceous; Figure 1). Flysch bedrock is formed in the foreland basins of developing mountain regions (syn-orogenic) and thus consists of a combination of deep sea deposits and weathering material from the forming Alps. Flysch bedrock is usually highly heterogeneous. The Schlieren Flysch consists of shales with calcareous phyllites (slate) and sandstone banks and covers the lower ~8 ha of the catchment (Figure 1b). The Ragazer Flysch consists of sheared and wrapped calcareous rocks with layers of marl, shales, and marly carbonates (South Helvetic) and covers the largest part of the catchment (~11 ha; Figure 1b). The south-eastern corner of the catchment (~1.5 ha) is underlain by Wild Flysch, which is similar to the Schlieren Flysch in lithology (shales with calcareous slates and banks of sandstone). The bedrock is assumed to be poorly permeable (Mohn, Schürmann, Hagedorn, Schleppi, & Bachofen, 2000). The calcareous nature of the bedrock results in predominantly calcium bicarbonate groundwater and streamwater with relatively high solute concentrations (range for groundwater and stream water during the campaigns presented in this study: 50–400 and 50–250 ppm, respectively; weighted average of 2 years of daily streamwater measurements at the neighbouring Erlenbach catchment: 103 ppm; Schleppi, Waldner, & Fritschi, 2006).

The landscape is influenced by landslides and soil creep, which have resulted in complex microtopography with steep slopes (more than 20°) and flatter areas (Figure 1a); the average slope is 20° (Rinderer et al., 2014). Landslides expose fresh sediment and bedrock, leading to more weathering and heterogeneous soils. The steep parts of the catchment and the ridge sites are covered by an open coniferous forest (~11 ha, *Picea abies* L. with an understory of *Vaccinium* sp.; Hagedorn, Schleppi, Waldner, & Flühner, 2000), whereas the flat and concave areas are characterized by moor landscapes or wet grasslands (~7 ha; Figure 1a). The grassland areas in the upper part of the catchment are used for cattle grazing during summer. Springs and stream heads emerge at the transitions from steep to flat areas. The streams have a step-pool morphology (Molnar, Densmore, McArdell, Turowski, & Burlando, 2010). Most of the streams are not deeply incised and have almost no riparian zone (Hagedorn, Schleppi, Waldner, & Flühler, 2000).

Soil depth ranges from 0.5 m at the ridge sites to more than 2.5 m in depressions and is weakly related to slope (van Meerveld et al., 2018). The soils consist mainly (>85%) of clay and silt (Schleppi et al., 1998). Mollic Gleysol, with a top soil that is high in carbonate and a permanently reduced Bg horizon, occur at the grassland and moor sites, where the water table is persistently close to the soil surface (Rinderer et al., 2014). The redox potentials of these soils are positive near the surface (up to +600 mV) and can decline down to −250 mV at larger depths (up to 50 cm; Hagedorn et al., 2000). The mostly forested ridge sites are characterized by an umbric Gleysol with an oxidized Bw horizon with macropores (Mohn et al., 2000; Schleppi et al., 1998) and have a high redox potential (+200 to +800 mV; Hagedorn et al., 2000).

2 | METHODS

2.1 | Hydrometric data

The research area consists of seven nested catchments (C1 to C7), increasing in size from C1 (~0.2 ha) to C7 (20 ha; Figure 1b). Stream stage was measured at the outlet of the catchments using H-flumes (C1 and C2), V-notch weirs (C3–C5), and at natural river cross-sections (C6 and C7). The rating curves to determine streamflow were obtained by salt-dilution gauging.

In each subcatchment, eight or nine groundwater wells were installed in 2010 according to the distribution of the TWI values (Rinderer et al., 2014). In total, the groundwater monitoring network consists of 51 wells (Figure 1b). The wells are numbered based on the rank of the TWI within each subcatchment. For example, C1.1 is the well with the lowest TWI in the smallest subcatchment, whereas C3.9 is the well with the highest TWI in one of the intermediately sized subcatchments. The wells were installed at the soil–bedrock interface (by manual augering) and are screened over their entire length, except for the upper 10 cm. The depth of the wells varies between 0.45 and 2.14 m.

Groundwater and streamwater levels were measured at 5-min intervals during spring to fall (May to November) and at 10-min intervals during winter using capacitance water level loggers (Odyssey Dataflow Systems Pty Limited) or pressure transducers (DCX-22 CTD, KELLER AG für Druckmesstechnik or STS DL/N 70, Sensor Technik Sirnach AG). The pressure measurements were corrected for changes in atmospheric pressure using barometric pressure data (10-min interval), corrected for temperature and elevation effects, from the MeteoSchweiz meteorological station in Einsiedeln (910 m above sea level). For six of the 51 sites, groundwater level data were available only for the summer period because the groundwater tubes were removed before winter to avoid damage of the tubes due to the overlaying snow pack. For more detailed information about the hydrometric monitoring network, see Rinderer et al. (2014, 2016).

Precipitation was measured at three locations (RG1–3 in Figure 1b) using 0.2-mm resolution tipping bucket raingauges (Odyssey Dataflow Systems Pty Limited). Because RG1 and RG2 had the most complete dataset, we used the average of RG1 and RG2 in our analyses, unless explicitly stated otherwise.

2.2 | Groundwater and streamflow chemistry and isotopic composition

Nine snapshot sampling campaigns (SC1–SC9) were carried out during different conditions between May and November of 2016 and 2017 (Table 1). Groundwater samples were collected from all wells with a sufficient groundwater volume, and streamwater samples were collected at the outlets of the seven subcatchments (Figure 1b). All wells were purged the day before sampling by pumping them dry or extracting at least two times the well volume. 1

The EC of all water samples was measured directly in the field using a Multi 3420 conductivity probe (WTW Measurement Systems Inc). The samples were collected in polyethylene bottles (50–300 ml) and stored at 4°C until preparation and analysis within a week after collection, except for SC8 (August 2017) for which the samples were frozen (−18°C) after collection until sample preparation a few days before analysis. Sample preparation involved filtering (0.45 µm, SimplepureTM Syringe Filter) and redistribution of the sampled water into three different vials to measure cation and anion concentrations and the isotopic composition. The samples for cation analysis were acidified with 50 µl of 50-mM HNO_3^- solution to mobilize trace metals and reduce precipitation and microbial activity. The cation and anion analyses were performed at the Physics of Environmental Systems laboratory at ETH Zürich, using an ion chromatograph (861 Advanced Compact IC, Metrohm) for anions and a mass spectrometer (ICP-MS 9700, Agilent Technologies) for cations, except for the first campaign (SC1), for which the samples were analysed on an emission spectrometer (ICP-OES 5100, Agilent Technologies). Calibration curves were obtained from measurements immediately before or after the sample analyses with five calibration standards. Stable water isotope analyses were performed with a Cavity Ring-Down Spectrometer (L2140-i or L2130-i Picarro Inc.) at the isotope laboratory of the Chairs of Hydrology at the University of Freiburg.

2.3 | Ion balance, transformation of the data, and temporal variability analyses

For the interpretations on the ion balance, the water samples were assumed to be electroneutral. We estimated the bicarbonate concentrations from the difference between the positive load (sodium, potassium, calcium, and magnesium) and the negative load (sulfate and chloride). This neglects the possible influence of dissolved organic carbon (DOC) on the electrical balance and is thus an upper boundary. We used the estimated bicarbonate concentrations only to describe the general composition of the water and did not use it for other analyses.

The normality of the frequency distributions of the concentrations, EC, δ^2H, and $\delta^{18}O$, was tested using the Shapiro–Wilk test of normality (Table 2). We used a p value of 0.05 for all statistical analyses. Only EC, δ^2H, $\delta^{18}O$, and the concentrations of calcium and lithium were approximately normally distributed; the data for all other elements were log-transformed (Table 2). We used Kruskal–Wallis one-way analysis of variance to test if the median concentrations were different for the geological units. Only the differences between the Schlieren Flysch and the Ragazer Flysch were considered because only two wells were located in the Wild Flysch.

The significance of the differences between the median concentrations for the different sampling campaigns was tested with a pairwise Wilcoxon test. Additionally, we calculated the temporal variability for each parameter as follows: We calculated the average concentration across the catchment for each campaign, selected the maximum average concentration, and subtracted the minimum average concentration, and subsequently, divided this range in average

TABLE 1 Overview of the conditions prior to the nine snapshot sampling campaigns: the streamflow at the outlet of C4 and C5 at 8:00 am on the day of sampling (Q_{C4} and Q_{C5}), the sum of precipitation in 7 days (P_7) and 2 days (P_2) prior to the sampling campaign, the average temperature in the week before the sampling campaign (T_7), and the number of wells that contained water and could thus be sampled

Sampling campaign	Date	Streamflow (L/s/km^2)	Streamflow (L/s/km^2)	Precipitation 7 days prior to sampling (mm)	Precipitation 2 days prior to sampling (mm)	Average temperature 7 days prior to sampling (°C)	Number of wells sampled
1	06/20/2016	38	60	134	27	11.3	42
2	07/19/2016	28	4	96	0	13.6	34
3	08/31/2016	5	<3	23	23	19.1	38
4	10/05/2016	3	4	11	0.6	12.2	37
5	10/31/2016	3	5	51	0.1	6.6	39
6	05/17/2017	95	24	37	0.3	12.8	36
7	06/20/2017	4	4	37	0	18.4	38
8	08/24/2017	—	5	16[a]	1.0[a]	16.0	35
9	10/12/2017	—	<3	49[a]	0.1[a]	8.3	35

Note. The streamflow of subcatchments C4 and C5 was used as a proxy for catchment wetness because these stream level data series are most complete and the rating curves are most reliable because of the V-notch weirs. Dashes (—) indicate no data.

[a]Average precipitation for RG1 and RG3 instead of RG1 and RG2 (see Figure 1b for the location of the raingauges).

TABLE 2 Overview of the mean, median, and range of the electrical conductivity (EC), isotopic composition, and ion concentrations (ordered by relative abundance) for all sampling campaigns (BDL = below detection limit), the median concentrations for the two geologic units and corresponding *p* values (Kruskal–Wallis one-way analysis of variance; significant differences are highlighted with an asterisk [*]), the transformation that was applied to the data before all analyses and corresponding *p* values of the normality test (Shapiro–Wilk) after the transformation, and the sampling campaigns for which the parameter was measured

		Entire catchment						Geologic dependence			Transformation		
Parameter and units		Mean	Median	Min.	Max.	Median spatial range	Median temporal range	Median Schlieren	Median Ragazer	*p* value geology	Transformation	*p* value normality	Sampling campaigns
EC	µS/cm	388	387	68	780	542	285	423	368	0.01*	None	4.22E-01	1–9
δ^2H	‰	−76.0	−76.8	−91.1	−47.1	18.5	18.1	−77.0	−76.6	0.200	None	1.28E-06	1–9
$\delta^{18}O$	‰	−11.0	−11.1	−12.9	−7.6	2.0	2.2	−11.1	−11.1	0.49	None	1.15E-06	1–9
D_{ex}	‰	12.0	12.0	7.5	14.6	2.9	1.6	11.8	12.2	0*	None	6.14E-05	1–9
Ca	mg/L	57.0	55.6	11.7	124.9	76.8	50.2	57.4	53.0	55.6	None	6.78E-03	1–9
SO_4	mg/L	3.6	2.6	BDL	39.6	18.6	4.4	2.7	2.5	2.6	Log	8.35E-06	1–9
Mg	mg/L	2.2	1.5	0.25	9.3	6.3	1.1	2.1	1.1	1.5	Log	2.79E-04	1–9
Na	mg/L	1.6	0.9	0.2	20.4	17.1	0.9	1.1	0.8	0.90	Log	1.17E-11	1–9
Cl	mg/L	0.83	0.70	BDL	13.0	2.8	1.7	0.50	0.76	0.70	Log	9.52E-08	1–9
Zn	mg/L	0.60	0.05	0.0004	18.5	3.1	0.93	0.03	0.11	0.05	Log	8.25E-03	1–9
Mn	mg/L	0.60	0.15	0.0001	8.9	4.3	0.43	0.30	0.09	0.15	Log	1.06E-04	1–9
K	mg/L	0.53	0.40	0.01	3.0	2.1	0.52	0.29	0.47	0.40	Log	4.94E-04	1–9
Fe	mg/L	0.38	0.01	BDL	11.4	3.9	0.19	0.02	0.01	0.01	Log	1.81E-12	1–9
Sr	mg/L	0.34	0.30	0.07	0.70	0.96	0.16	0.36	0.25	0.30	Log	8.11E-01	6–9
NO_3	mg/L	0.28	0.10	BDL	0.80	1.4	0.67	0.04	0.10	0.10	Log	3.01E-03	1–9
Ba	µg/L	99	53.2	5.2	1984	315	87.1	57.2	42.0	0.002*	Log	3.68E-06	2–9
F	µg/L	77	60	BDL	750	270	134.5	50	60	0.04*	Log	3.76E-04	1–9
Cu	µg/L	65	7.0	BDL	1044	466	26.7	7.30	7.60	0.1	Log	1.27E-08	1–9
PO_3	µg/L	17	0	BDL	390	1	85	0	0	0.52	Log	1.40E-04	1–9
B	µg/L	10	8.65	0.27	47.8	25.9	10.2	9.20	8.60	0.09	Log	1.45E-06	2–9
Sn	µg/L	9.30	8.04	1.32	22.1	–	20.7	11.1	7.70	0.07	Log	4.60E-03	1
Al	µg/L	5.70	0.01	BDL	74.5	0.03	14.1	0.01	0.01	0.72	Log	3.73E-11	1, 6, 7, and 9
Ni	µg/L	3.00	1.90	BDL	43.8	15.1	4.70	1.90	1.90	0.20	Log	5.14E-03	1–9
Pb	µg/L	2.60	0.40	BDL	45.0	23.6	3.10	0.30	0.55	0.03*	Log	8.10E-03	1–9
Li	µg/L	1.70	1.05	BDL	5.70	0.94	1.74	0.80	1.70	0.16	None	8.46E-06	1–5
Co	µg/L	0.79	0.38	0.01	9.40	3.70	0.96	0.60	0.30	0.01*	Log	5.09E-03	2–9
As	µg/L	0.40	0.28	0.05	2.50	1.51	0.35	0.30	0.30	0.51	Log	8.90E-02	6–9
Mo	µg/L	0.13	0.09	0.012	0.67	0.56	0.11	0.10	0.06	<0.001*	Log	1.18E-01	6–9

(Continues)

TABLE 2 (Continued)

Parameter and units		Entire catchment						Geologic dependence			Transformation		
	Mean	Median	Min.	Max.	Median spatial range	Median temporal range	Median Schlieren	Median Ragazer	*p* value geology	Transformation	*p* value normality	Sampling campaigns
V	µg/L	0.09	0.06	0.01	0.62	0.37	0.07	0.05	0.08	0.008*	Log	3.64E-02	6–9
Cr	µg/L	0.09	0.07	0.006	0.59	0.27	0.09	0.07	0.07	0.81	Log	6.54E-02	1–9
Se	µg/L	0.07	0.05	0.01	0.39	0.26	0.04	0.05	0.05	0.78	Log	1.04E-02	1–9
Cd	µg/L	0.05	0.03	BDL	0.96	0.23	0.07	0.03	0.03	0.10	Log	6.11E-04	2–9
Tl	µg/L	0.01	0.01	BDL	0.08	0.02	0.02	0.01	0.01	0.07	Log	6.70E-09	4–9
Bi	µg/L	0.01	0	BDL	0.24	0.03	0.001	0	0.01	0.37	Log	3.49E-07	4–7 and 9
Be	µg/L	0.005	0.003	BDL	0.03	0.02	0.004	0.005	0.003	0.08	Log	8.77E-03	6–9

Note. The spatial range is the difference between the minimum and maximum concentrations measured during a sampling campaign, whereas the temporal range is the difference between the minimum and maximum concentrations measured at a single well. D_{ex} is the deuterium excess, calculated as $D_{ex} = \delta^2H - 8*\delta^{18}O$.

concentrations by the average of all groundwater concentrations measured during the study.

2.4 | Hydrodynamic and topographic site attributes

We hypothesized that the solute concentrations at a given location are influenced by the local water flux and can thus be described by hydrodynamic or topographic characteristics. Hence, we determined several site characteristics for each well in order to explore the relation between these characteristics and the spatial variability in groundwater chemistry. The characteristics included well depth, geology, variability of the groundwater levels (hydrodynamic attributes), and topographic attributes (Table 3).

The hydrodynamic attributes describe the wetness of the site and the frequency and duration that the water table was in the more conductive upper soil horizons. The hydrodynamic attributes were calculated for each sampling location based on the groundwater levels measured between December 1, 2015, and November 29, 2017, aggregated to hourly values to reduce noise in the data. Depending on whether the measurements were taken throughout the year or only during the summer period, the groundwater level time series length varied between 13.5 and 36.4 months. The hydrodynamic attributes included the median depth of the groundwater level below the surface (median), skewness of the frequency distribution of the groundwater levels (skew), the fraction of time that the groundwater level was within 10 or 30 cm from the land surface (persistence, PST10, and PST30, respectively), and the flushing frequency (FF30), defined as the number of times that the groundwater level rose into the more conductive part of the soil (set to 30 cm below the soil surface) divided by the time series length.

The topographic characteristics describe the accumulation of water and solutes from upslope and/or the drainage. Rinderer et al. (2014) showed that the spatial variability in groundwater level dynamics can be described by surface topographic indices. We therefore assumed that they might also describe the spatial variability in groundwater chemistry. The topographic attributes were calculated based on a 6 × 6 m digital terrain model derived from LiDAR data. The 6-m resolution is consistent with previous analyses on the relation between topography on groundwater level dynamics (Rinderer et al., 2014, 2016, 2017) and is considered the optimum resolution for this catchment. We selected the site-specific topographic attributes that are significantly correlated with the groundwater level dynamics (Rinderer et al., 2014): slope, TWI (Beven & Kirkby, 1979), accumulated area (c.f. Seibert & McGlynn, 2007), as well as several other attributes that are commonly used in the literature to describe topographic characteristics: profile curvature, plan curvature, downslope index (Hjerdt, 2004), gradient to creek (the gradient from a well to the nearest stream cell), and elevation above the nearest stream along the D8-flow path. We furthermore included attributes that describe the upslope flow pathways because they may influence the solute load or flushing frequency: average flowpath length between the well and the divide, maximum upslope flowpath length, and average slope of the upslope flowpath. All topographic attributes were

TABLE 3 Overview of the hydrodynamic and topographic attributes, their corresponding abbreviations (Abbr.), transformation used for normalization of the data, and *p* value of the normality test (Shapiro–Wilk) after transformation

Attribute	Description	Abbr.	Transformation	*p* value normality
Hydrodynamic				
Flushing frequency	Number of times the groundwater level rose into the more conductive part of the soil (set to 30 cm)	FF30	Log	4.3E-3
Median	Median groundwater level	Median	None	9.4E-2
Persistence 10	Fraction of time the groundwater level was within 10 cm from the surface	PST10	Log	2.4E-2
Persistence 30	Fraction of time the groundwater level was within 30 cm from the surface	PST30	Log	7.0E-6
Skewness	Skewness of the groundwater level distribution	Skew	None	1.5E-6
Topographic				
Average flowpath length	Average length of the flowpath from the site to the nearest stream cell	AFL	Log	7.5E-2
Average flowpath slope	Average slope of the flowpath from the site to the nearest stream cell	AFS	None	2.2E-7
Downslope index	Horizontal distance water would have to travel along a flowpath to drop a given vertical distance (set to 6 m)	DI	None	3.4E-1
Elevation above stream	Gravity potential relative to the nearest stream cell	EAS	Log	1.6E-3
Accumulated area	Accumulated area draining into the grid cell	AA	Log	1.2E-5
Gradient to creek	The gradient of the groundwater flow path (gravity potential divided by distance to the nearest stream cell)	GTC	Log	1.1E-1
Max length flowpath	Maximum length of the upslope flowpath	MUFL	Log	8.1E-2
Slope	Local slope	SL	None	6.0E-2
Plan curvature	Curvature parallel to the slope direction	PLC	Log	3.7E-4
Profile curvature	Curvature perpendicular to the slope direction	PRC	None	2.8E-1
Topographic wetness index	Wetness index based on the accumulated area and local slope	TWI	Log	1.4E-1

computed using the Whitebox Geospatial Analysis Tools (Lindsay, 2016). To identify the stream cells, we used a channelization threshold of 25 cells and verified the output with field observations. All attribute values were log-normal transformed prior to the analyses, except for the median groundwater level, the skewness of the distribution of the groundwater level, the slope, the profile curvature, and the average flowpath slope (Table 3).

2.5 | Correlation between site attributes and groundwater composition

The relation between hydrodynamic and topographic attributes and groundwater chemistry was explored for the individual elements and the overall water composition. First, we used Spearman's rank correlation analysis to identify which hydrodynamic and topographic attributes best describe the variability in the concentrations of the different elements. Next, we used a random forest (RF) analysis to quantify the relative importance of site attributes to describe the concentrations of calcium, sulfate, copper, and manganese, representing the major cations, anions, heavy metals, and transition metals, respectively. Finally, we use a principal component analysis (PCA) for the element concentrations and compared the principal components with the site attributes that were most important in the RF. This allowed for the comparison of the overall water composition to the relevant hydrodynamic and topographic attributes. The isotopic composition was not used in this PCA due to the seasonal changes in the isotopic composition of precipitation and thus groundwater. We used normalized data for the Spearman's rank correlation and RF analysis and additionally scaled (z-transformation) the data for the PCA. We found that using Flysch lithology as another possible site attribute did not change the results; we thus neglected the different Flysch lithologies in these analyses.

2.6 | Delineation of groundwater types and comparison to streamflow

The persistence of the spatial variability of the concentrations was assessed by calculating the mean relative difference (M_{RD}) for each parameter (concentrations, EC, and isotopic composition) and each well using Equations (1) and (2), where x_j is the measured value for that well during sampling campaign j, μ_j the average value for all wells sampled during sampling campaign j, σ the standard deviation for sampling campaign j, and n_{sc} is the number of sampling campaigns for which there are samples for that well.

$$R_{Dj} = \frac{x_j - \mu_j}{\sigma_j}. \quad (1)$$

$$M_{RD} = \frac{1}{n_{sc}} \sum_{j=1}^{n_{sc}} R_{Dj}. \quad (2)$$

We identified different shallow groundwater types based on *k*-means clustering on the M_{RD} values for the different parameters. Based on the Calinski–Harabasz (CH) criterion, we obtained between three and five clusters. We decided on four clusters because of the relatively even distribution of wells per cluster (14, 11, nine, and seven wells, respectively), even though the Calinski–Harabasz value for five clusters was slightly lower. The significance of the differences in the median concentrations and the hydrodynamic and topographic attribute values for the different clusters (i.e., groundwater types) was tested with a Kruskal–Wallis one-way analysis of variance.

PCA was used to compare the streamflow chemistry (element concentrations and isotopic compositions) with the chemistry of the shallow groundwater. We performed the PCA for each campaign separately to avoid comparison of streamwater of one campaign to groundwater sampled during another campaign.

3 | RESULTS

3.1 | Hydrochemical setting

The most abundant solutes (>100 μg/L) in the shallow groundwater were calcium, bicarbonate, sulfate, magnesium, sodium, chloride, potassium, strontium, and manganese. The chemistry of the shallow groundwater was dominated by calcium; its relative contribution to the electrical balance was larger than 80% for 94% of the samples (median: 93%). For the samples for which calcium was less dominant for the electrical balance, the importance of magnesium was higher. For 95% of the samples, the contribution of magnesium to the electrical balance was less than 5%, although for some samples (all from specific wells), the contribution was as high as 34%. Estimates of bicarbonate concentrations suggest that it is the dominant anion for all samples, whereas sulfate was the second most abundant anion (Table 2).

For roughly half of the analysed parameters, there was a significant difference in the median concentration between the Ragazer Flysch and the Schlieren Flysch (Table 2). EC and concentrations of magnesium, sodium, manganese, cobalt, and molybdenum were significantly higher in the lower part of the Studibach catchment (Ragazer Flysch), whereas concentrations of chloride, potassium, zinc, nitrate, and vanadium were significantly higher in the upper part of the catchment (Schlieren Flysch).

3.2 | Temporal and spatial variability in groundwater chemistry

The shallow groundwater chemistry varied temporally, with solute concentrations increasing during summer when water levels and catchment wetness decreased (Figure 2). Although groundwater EC increased during the summer, there were only two significantly different groups of sampling campaigns: the two campaigns directly after

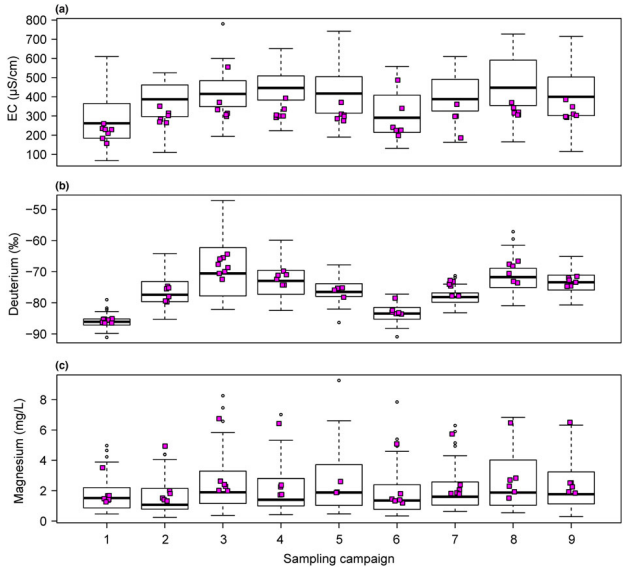

FIGURE 2 Boxplots of (a) electrical conductivity (EC, μS/cm), (b) deuterium (‰), and (c) magnesium concentrations (mg/L) in shallow groundwater and the concentrations in streamwater (pink squares) for all nine sampling campaigns (SC1 to SC9). For information on the conditions during the different sampling campaigns, see Table 1. The box represents the 25th–75th percentile, the thick line represents the 50th percentile, the whiskers extend to the 25th percentile − 1.5*interquartile range and 75th percentile + 1.5*interquartile range, and the dots represent the outliers

snowmelt (SC1 in June 2016 and SC6 in May 2017) and all other campaigns (Figure 2). The isotopic composition was more temporally variable than the other hydrochemical parameters. There were three significantly different groups: early summer directly after snowmelt, which was characterized by isotopically more depleted water (SC1 and SC6), late summer, which was dominated by isotopically more enriched water (SC3, SC4, and SC8), and a mixture of both (SC2, SC5, SC7, and SC9). The temporal variability of the parameters across the catchment was largest for lead, aluminium, and cobalt (3.5, 2.4, and 1.7, respectively) and smallest for beryllium, magnesium, and sodium (0.03, 0.06, and 0.07, respectively).

The spatial variability in shallow groundwater chemistry was large, with concentration ranges spanning up to five orders of magnitude for elements such as calcium, manganese, and zinc. The median spatial variability (i.e., the median difference between the minimum and maximum concentrations measured during a single sampling campaign; n = 9) was larger than the median temporal variability (i.e., the median difference between the minimum and maximum concentrations measured for a single well; n = 47) for almost all parameters (Table 2). Exceptions were phosphate and aluminium, for which the temporal variability was larger than the spatial variability (but this could be influenced by the large number of samples with concentrations below the detection limit), and the stable water isotopes, for which the spatial variability was as large as the temporal variability.

3.3 | Correlation between site attributes and groundwater composition

The extent to which the spatial variability in chemistry could be explained by the hydrodynamic and topographic attributes depended on the element of interest (shown for M_{RD} in Table 4). Spearman's rank correlations ranged from ~0.01 to 0.67 when all campaigns were analysed together and were generally higher for the individual sampling campaigns (range between ~0.005 and 0.81). The heavy metal (copper, zinc, nickel, and lead), transition metal (iron, manganese, and cobalt), and potassium concentrations were relatively well correlated with the topographic and hydrodynamic attributes. The heavy metal concentrations were higher at sites that were predominantly dry, whereas the transition–metal concentrations were usually higher at sites that were predominantly wet. Potassium concentrations were strongly correlated with multiple hydrodynamic and topographic attributes and were higher at drier sites.

Although the correlations for the other solute concentrations, isotopic composition, and EC with the topographic and hydrodynamic attributes were weak, they sometimes suggested an upper boundary. For instance, at locations with a relatively low downslope index, indicating flatter areas downslope and therefore areas with likely more stagnant groundwater, groundwater EC values were up to 780 μS/cm, whereas at steeper sites, where one would expect better drainage and faster groundwater flow, the maximum EC was ~450 μS/cm. At drier sites (i.e., lower median groundwater levels), the groundwater was on average more isotopically depleted in June 2016, directly after the snowmelt period, compared with groundwater at wet sites (i.e., shallower groundwater levels). This could indicate that snowmelt had a more pronounced influence on the groundwater isotopic composition at drier sites and, likewise, that the composition at these sites might be more strongly influenced by recent inputs (such as snowmelt or precipitation).

The RF results suggest that calcium concentrations were best explained by a combination of topographic attributes (gradient to creek, downslope index, plan curvature, and accumulated area; Figure 3a) but that the hydrodynamic attributes describing the shape and magnitude of the groundwater levels (skew and median) were most powerful in explaining the spatial variability in sulfate concentrations (Figure 3b). There was a more gradual decrease in the explanatory power of the hydrodynamic and topographic attributes for copper and manganese (Figure 3c,d).

The two components of the PCA for only the most abundant solutes except bicarbonate (>100 μg/L: calcium, sulfate, magnesium, sodium, chloride, potassium, strontium, and manganese) explained 50% of the variability in the 318 shallow groundwater samples (PC1: 29% and PC2: 21%). PC1 was most strongly correlated with the concentrations of magnesium ($\cos^2$ = 0.70), calcium ($\cos^2$ = 0.57), and sodium ($\cos^2$ = 0.42), whereas PC2 was most influenced by the concentrations of chloride ($\cos^2$ = 0.51) and sulfate ($\cos^2$ = 0.44). When all elements were included in the PCA, the explanatory power decreased to 33% (PC1: 18% and PC2: 15%; Figure 4a). PC1 was most strongly determined by copper, lead, cadmium, and zinc ($\cos^2$ = 0.74, 0.59, 0.49, and 0.45, respectively) and PC2 by calcium, barium, and strontium ($\cos^2$ = 0.57, 0.44, and 0.41 respectively). The (base) cations (Ca, Mg, Na, Ba, and B), heavy metals (Cu, Pb, and Zn), and transition metals (Mn, Fe, and Co) grouped together in the principal component space for all campaigns, except SC8 (August 2017), but the grouping was less pronounced when all campaigns were analysed together (Figure 4a).

When all solutes were used in the PCA, PC1 was strongly correlated with the hydrodynamic and topographic attributes but PC2 was not (Figure 4b). PC1 was strongly negative correlated with the persistence of the water table in the upper 10 and 30 cm of the soil (PST10: r^2 = −0.55; PST30: r^2 = −0.65) and the median groundwater level (r^2 = −0.64), and strongly positive correlated with slope (r^2 = 0.52) and ruggedness (r^2 = 0.50). The correlations of PC2 with the attributes were very low (which corroborates the findings of the Spearman's rank analysis), with a maximum of 0.16 for TWI.

3.4 | Delineation of groundwater types based on the mean relative difference (M_{RD})

Clustering based on the average spatial variability, expressed as M_{RD}, yielded four groundwater types (Figure 5). For each element, the median concentrations were significantly different for at least one of the groundwater types, except for nitrate, aluminium, arsenic, selenium, bismuth, beryllium, and tin. The wells with water Type I were characterized by above average concentrations of transition metals, and below-average concentrations of heavy metals, trace metals, calcium, and EC. These wells were mostly located in flat, wet areas,

TABLE 4 Spearman's rank correlation coefficients (r_s) of the relationships between the mean relative difference (M_{RD}; Equation 2), and the hydrodynamic and topographic attributes (for abbreviations, see Table 3)

	Hydrodynamic attributes				Topographic attributes					
Parameters	Median	Skew	PST10	PST30	DI	EAS	GTC	PLC	SL	TWI
EC	0.17	−0.10	0.04	0.15	0.07	0.28	0.02	−0.04	−0.12	−0.08
$\delta^{18}O$	−0.09	0.02	−0.11	0.00	−0.18	0.18	0.15	−0.10	0.17	−0.11
$\delta^{2}H$	−0.09	0.02	−0.06	0.00	−0.14	0.15	0.15	−0.01	0.17	−0.02
D_{ex}	0.09	−0.16	0.21	0.10	0.19	−0.23	0.01	0.23	−0.06	0.38
Ca	−0.01	0.02	−0.11	−0.01	−0.06	0.32	0.12	−0.19	0.06	−0.23
SO4	−0.36	0.49	−0.44	−0.31	−0.38	0.22	0.27	−0.25	0.27	−0.29
Mg	0.06	0.14	−0.08	0.06	−0.07	0.09	−0.03	−0.02	−0.06	−0.25
K	−0.64	0.44	−0.68	−0.65	−0.51	0.36	0.33	−0.27	0.46	−0.29
Na	0.05	0.15	0.06	0.11	−0.07	−0.12	−0.08	−0.02	−0.04	−0.10
Cl	−0.55	0.51	−0.51	−0.54	−0.53	0.10	0.35	−0.41	0.48	−0.31
Mn	0.76	−0.61	0.71	0.73	0.64	−0.33	−0.53	0.44	−0.59	0.36
Fe	0.49	−0.56	0.59	0.50	0.56	−0.47	−0.58	0.36	−0.57	0.35
Co	0.64	−0.57	0.65	0.59	0.59	−0.22	−0.41	0.35	−0.52	0.26
F	−0.15	0.19	−0.21	−0.15	−0.13	0.06	−0.01	−0.01	0.10	−0.18
NO_3	−0.52	0.48	−0.47	−0.53	−0.47	0.22	0.49	−0.32	0.39	−0.19
PO_3	−0.52	0.51	−0.40	−0.49	−0.44	0.42	0.35	−0.25	0.48	−0.39
Al	−0.33	0.14	−0.37	−0.33	−0.29	0.09	−0.02	−0.19	0.25	−0.05
B	0.18	−0.01	0.09	0.19	0.07	0.06	−0.14	0.13	−0.12	−0.05
V	−0.60	0.46	−0.57	−0.56	−0.57	0.16	0.29	−0.38	0.52	−0.12
Cr	−0.57	0.26	−0.51	−0.60	−0.37	0.15	0.20	−0.46	0.33	−0.21
Ni	−0.57	0.34	−0.53	−0.61	−0.41	0.39	0.37	−0.49	0.49	−0.32
Cu	−0.71	0.54	−0.66	−0.71	−0.72	0.45	0.53	−0.53	0.68	−0.40
Zn	−0.56	0.32	−0.47	−0.57	−0.44	0.51	0.42	−0.35	0.50	−0.33
As	0.33	−0.17	0.25	0.34	0.21	−0.22	−0.30	0.14	−0.26	0.08
Se	−0.42	0.21	−0.45	−0.49	−0.23	0.06	0.02	−0.25	0.11	−0.09
Sr	0.17	−0.15	0.12	0.11	0.10	0.21	0.02	0.19	−0.15	−0.03
Mo	0.19	−0.27	0.20	0.15	0.35	−0.10	−0.15	0.15	−0.38	−0.09
Cd	−0.42	0.13	−0.36	−0.41	−0.25	0.35	0.17	−0.22	0.26	−0.09
Ba	−0.10	0.01	−0.13	−0.16	−0.05	0.25	0.20	−0.20	0.06	−0.24
Tl	−0.43	0.34	−0.50	−0.47	−0.33	0.16	0.25	−0.28	0.38	−0.40
Pb	−0.68	0.45	−0.65	−0.69	−0.60	0.44	0.43	−0.43	0.59	−0.35
Bi	−0.65	0.21	−0.43	−0.64	−0.54	0.37	0.46	−0.36	0.54	−0.29
Be	−0.10	−0.09	−0.03	−0.10	−0.02	−0.27	−0.37	−0.04	−0.07	0.15
Li	−0.07	0.13	−0.14	−0.04	−0.25	0.50	0.43	−0.37	0.21	−0.49
Sn	−0.35	0.32	−0.54	−0.33	−0.44	0.05	0.05	−0.32	0.23	−0.10

Note. Cells are colour coded to more easily see the differences in the correlations (dark blue: $r_s \geq 0.5$; light blue: $0.2 \leq r_s < 0.5$; white: $-0.2 \leq r_s < 0.2$; light red: $-0.2 \leq r_s < -0.2$; dark red: $r_s \leq -0.5$). Elements are ordered by their grouping in the M_{RD}-cluster analysis.

where water tables were persistently close to the surface (Table 5). The wells with water Type II were characterized by relatively high concentrations of heavy metals, and some had an enriched isotopic signature compared with the catchment average. These wells were predominantly located in the upper parts of the different sub-catchments, on steep slopes, and had water levels that were significantly more variable (Table 5). The wells with water Type III were relatively deep wells (110 ± 30 cm) and except for two wells were all located in the lower part of the catchment. These wells had above average concentrations of magnesium and sulfate, low concentrations of transition metals, and a depleted isotopic signature compared with the other water types. The wells with water Type IV were

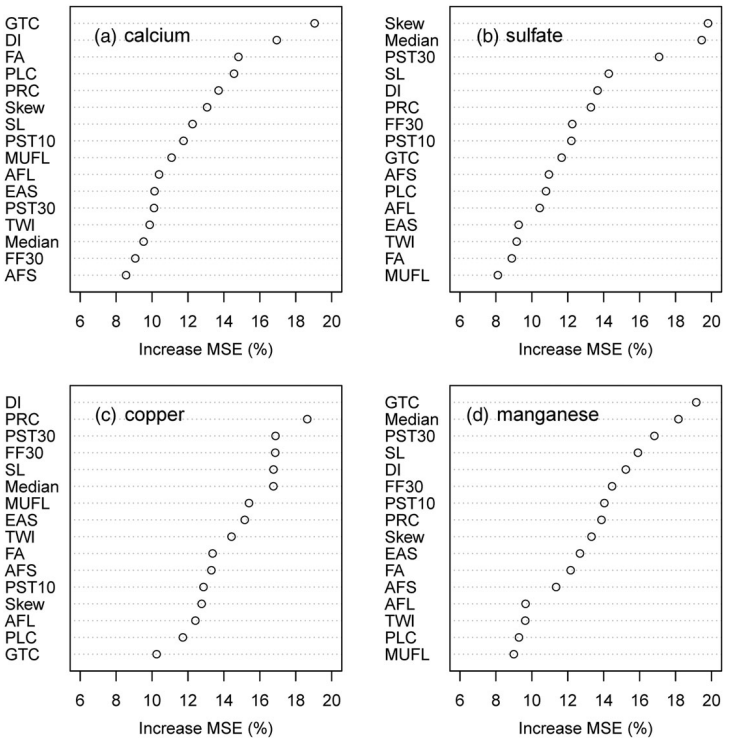

FIGURE 3 Predictive power of the hydrodynamic and topographic attributes (see Table 2 for attribute abbreviations) as represented by the percentage increase of the mean standard error (MSE, %) for (a) calcium, (b) sulfate, (c) copper, and (d) manganese

characterized by high magnesium and sulfate concentrations and relatively low concentrations of transition metals and heavy metals. They were located on the lower part of the hillslopes, predominantly in the lower part of the catchment.

3.5 | Comparison of groundwater and baseflow

The EC, $\delta^{18}O$, $\delta^{2}H$, and element concentrations in streamwater were within the range of the shallow groundwater, but the median concentrations were different for some elements. The median concentrations of manganese and zinc in streamwater were significantly lower than the median concentrations in groundwater for all nine sampling campaigns. The concentrations of chloride, iron, and fluoride were significantly lower in streamflow than in groundwater during one campaign (SC5, SC8, and SC3, respectively). Magnesium concentrations (Figure 2c) were relatively high in streamwater (7.1–8.8 and 5.5–9.1 mg/L in streamwater and groundwater, respectively). Streamwater EC was less variable in space and time (standard deviation for all streamwater and all groundwater samples: 69 and 138 μS/cm, respectively); the range in streamwater EC (157–555 μS/cm) was smaller than the range in groundwater EC (68–780 μS/cm; Figure 2a). The difference in the median EC for streamflow and groundwater was significant for the late summer and fall sampling campaigns (SC4, SC5, SC7, and SC8). The streamwater isotopic signature was most variable across space in SC3. The streamwater isotopic signature was most similar to the median groundwater isotopic composition in SC1 and SC6, when the range of $\delta^{18}O$ and $\delta^{2}H$ in groundwater and streamwater was smallest (Figure 2b). In SC4 and SC7 (late summer campaigns), streamwater was isotopically enriched compared with the median groundwater isotopic composition, but the difference was only significant for SC7.

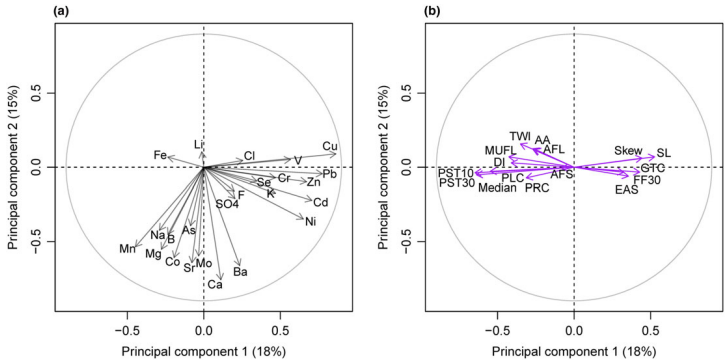

FIGURE 4 (a) Correlation of the element concentrations with the principal component axes (black arrows), and (b) correlation of the hydrodynamic and topographic attributes to the principal component axes (purple arrows) in the two-dimensional principal component space for all groundwater samples (n = 318). The length of the arrow represents the strength of the correlation. See Table 3 for the explanations of the abbreviations of the topographic and hydrodynamic attributes

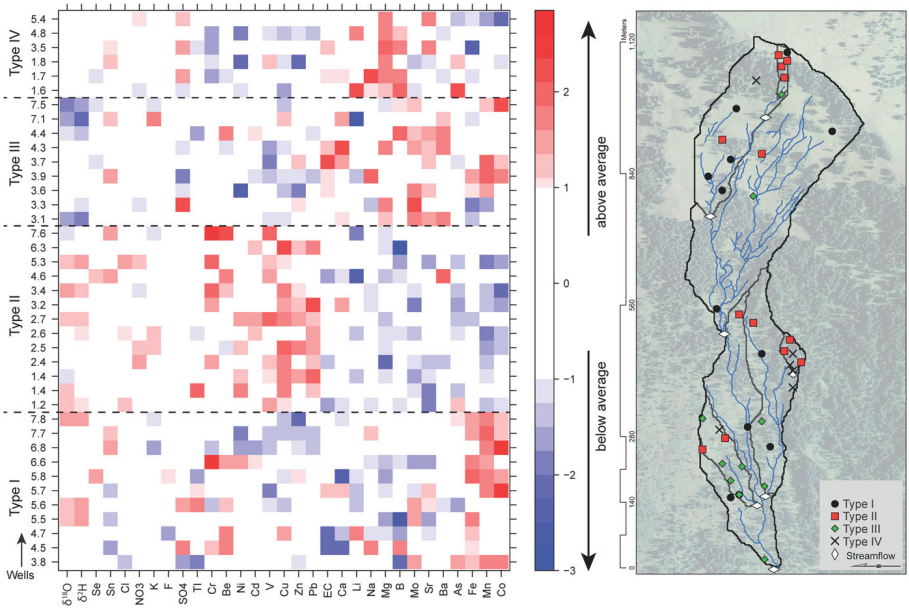

FIGURE 5 (left) Mean relative differences (M_{RD}; Equation 2) from the catchment average concentrations in groundwater for each well (rows with well number) and parameter (columns). Solutes and wells are ordered according to the four groundwater types (k-means clustering) based on the M_{RD} values; (right) map with wells colour coded according to the groundwater types (black circles = Type I, red squares = Type II, green diamonds = Type III, and black cross = Type IV)

TABLE 5 Median values for the hydrodynamic and topographic attributes that were significantly different between the groundwater types, and the Kruskal–Wallis *p* values comparing the median attribute values for each groundwater type

	Hydrodynamic attributes				Topographic attributes					
Groundwater type	Median (cm)	Skew (–)	PST30 (–)	PST10 (–)	DI (m)	TWI (–)	GTC (m)	PLC (–)	PRC (–)	SL (°)
I	−9.7[a]	−1.9[a]	0.85[a]	0.51[a]	29.0[a]	6.6[a]	9.6[a]	4.9[a]	0.3[a]	14.8[a]
II	−50.8[b]	0.4[b]	0.06[b]	0[b]	14.5[b]	4.4[ab]	26.1[b]	−0.9[b]	−0.8[b]	25.5[b]
III	−27.4[ab]	−0.6[a]	0.59[ab]	0.05[ab]	23.0[a]	4.3[b]	15.1[ab]	2.9[ab]	0.9[ab]	15.6[a]
IV	−24.7[ab]	−0.4[ab]	0.73[ab]	0.01[ab]	19.6[ab]	4.6[ab]	20.2[ab]	2.4[ab]	0.5[ab]	21.0[ab]
p value	.006	.005	.012	.003	<.001	0.049	.015	.003	.019	.001

Note. For attribute abbreviations, see Table 3. Different letters indicate significant differences in the median values of an attribute.

The streamflow samples plotted near the groundwater samples in the principal component space (shown for SC6 and SC3 in Figure 6 and Table 6). The streamwater did not exclusively resemble groundwater at spatial proximity or near to the stream or groundwater that was sampled upslope from the streamflow sampling site. The distance between the median streamflow coordinates and the median coordinates for the four groundwater types (derived from the clustering based on the mean relative differences) were smallest for groundwater Types I and II (Table 6). Streamflow was most similar to groundwater Type II (i.e., wells in steeper drier areas) in SC1 and SC6 and most similar to groundwater Type I, that is, the flatter, wetter sites, for all other campaigns. However, streamwater at C1 was more similar to groundwater Type IV (in SC1 and SC6) and groundwater Type III (in all remaining campaigns). In SC6, the streamwater at C4 was also very similar to groundwater Type III. Streamwater at C4 and C7 had a more enriched isotopic composition and slightly higher concentrations of copper and lead compared with the sites draining the northern part of the catchment during SC2 and SC3 (July and August 2016). For SC7 (June 2017), the downstream catchments (C3, C4, and C7) had a more enriched isotopic signature than the streams draining the upper part (C5 and C6).

4 | DISCUSSION

4.1 | Spatial variability in shallow groundwater chemistry

The nine snapshot sampling campaigns between early summer and fall in 2016 and 2017 highlight the large spatial variability in shallow groundwater chemistry across the 20-ha Studibach catchment during baseflow conditions. The concentrations of individual elements varied over at least two orders of magnitude and more than five orders of magnitude for some elements, such as iron, copper, and lead. The spatial variability was larger than the temporal variability for almost all parameters, which supports earlier findings of large spatial variability in subsurface water chemistry by Kendall et al. (2001) and Walker et al. (2003). The spatial variability in groundwater chemistry can be especially large in headwater catchments because flowpath lengths are short and internal mixing is limited, which increases the importance of physical and mineralogical heterogeneities (Walker et al., 2003). Differences in subsurface flow velocities, macropore flow, and flushing frequency due to landscape position have been suggested as the main hydrological controls on variations in subsurface water quantity and quality (Burns et al., 1998; Kendall et al., 2001; Welsch et al., 2001). At the Studibach catchment, the high heterogeneity of the Flysch bedrock material may have added to the large variability in shallow groundwater chemistry as well.

Similar to many other catchments (Anderson & Burt, 1978; Detty & McGuire, 2010; Jencso et al., 2009; Penna, Mantese, Hopp, Dalla Fontana, & Borga, 2015), topography explains a large part of the spatial variability in groundwater levels in the Studibach catchment (Rinderer et al., 2014). We hypothesized that topographic and hydrometric attributes would be able to explain the concentrations of major elements or the isotopic composition of groundwater in the Studibach catchment as well, but correlations between solute concentrations and topographic attributes (downslope index and slope) or hydrodynamic attributes (median groundwater level and water level persistence) were weak for most elements. Previous studies showed that concentrations of biogeochemical tracers, such as nitrate or DOC (Ogawa, Shibata, Suzuki, Mitchell, & Ikegami, 2006; Welsch et al., 2001) were correlated with surface topography. However, in the Panola Research Watershed where hillslope groundwater responses are short-lived and subsurface stormflow is related to the topography of the soil–bedrock interface rather than the surface topography, base cation concentrations were correlated with the bedrock accumulated area (Burns et al., 1998). Different trench sections also received unequal amounts of pipe flow, which may have affected the concentrations of base cations as well (Burns et al., 1998).

4.2 | Importance of landscape elements

Even though the correlations between individual topographic or hydrometric attributes and element concentrations were weak, we could distinguish four shallow groundwater types based on the mean relative differences in the concentrations from the catchment average (Figure 5). These types are related to three main hydrogeomorphic units: riparian-like zones or topographic hollows, which are located in

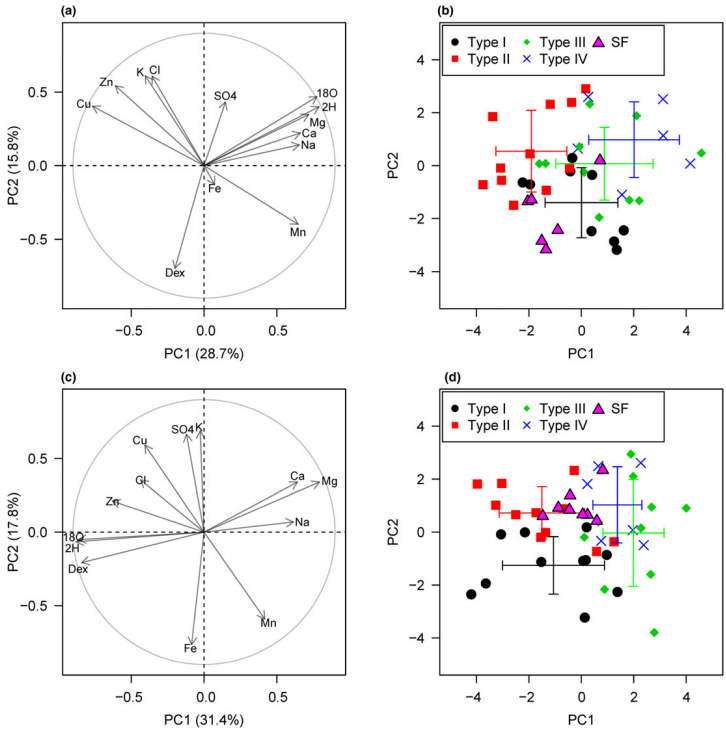

FIGURE 6 Two-dimensional PCA results for groundwater chemistry during (a, b) sampling campaign 6 (SC6; May 2017) and (c, d) sampling campaign 3 (SC3; August 2016). (a, c) The correlation of the element concentrations to the principal component axes, with the explanatory power of the individual principal components indicated at the axes. The length of the arrow represents the strength of the correlation. (b, d) The projection of the streamflow (SF, purple triangles) and groundwater samples coloured by groundwater type based on the clustering of the M_{RD} values (black circles = M_{RD} Type I, red squares = Type II, green diamonds = Type III, and blue crosses = Type IV, right panel) in the same two-dimensional principal component space. SC6 (c, d) is representative for early summer (after-melt) conditions, whereas SC3 (a, b) is representative for mid-summer conditions

flatter areas and have a high TWI and persistently shallow groundwater tables; hillslope and ridge sites that are located in steeper, well-drained areas; and "deeper" groundwater flow in areas that are relatively flat but have a lower TWI than the riparian-like areas and mostly (relatively) deep soils.

The sites that had water Type I one were characterized by low concentrations of heavy metals and trace metals and high concentrations of transition metals, a composition that is more commonly found in riparian areas (Hill, 2000). However, the Type I sites were not constrained to the areas adjacent to the streams but were rather spread throughout the catchment. Riparian zones in the Studibach catchment tend to be very narrow due to the steep topography of the catchment (Hagedorn et al., 2001), but areas with persistent shallow groundwater levels (Brinson, 1993) are common throughout the flatter areas of the catchment. The persistent high water levels close to the surface create almost permanent oxygen-reducing conditions (Megonigal, Patrick, & Faulkner, 1993), sustaining high concentrations of transition metals, such as manganese and iron. The adsorption of heavy metals on manganese and iron oxides can consequentially impact the heavy metal concentrations in the shallow groundwater (McKenzie, 1980), which were particularly low in these areas. The slightly enriched isotopic composition of the groundwater (compared with the catchment average) suggests that the shallow groundwater levels consist of a relatively large fraction of recent (i.e., summer) precipitation.

TABLE 6 Distances between the coordinates of the baseflow compositions and the median M_{RD}-based cluster coordinates (groundwater Types I–IV) in the principal component space (–) for the individual subcatchments (C1 and C3–C7; C2 was dry) and the median streamflow composition (median)

Sampling campaign	Catchment	Type I	Type II	Type III	Type IV
SC6	C1	1.43	2.32	1.42	0.99
	C3	2.52	3.90	2.37	4.92
	C4	1.66	3.24	0.74	4.07
	C5	2.36	3.57	3.44	4.75
	C6	2.13	2.14	2.60	4.11
	C7	1.98	2.05	3.22	3.95
	Median	2.17	1.88	2.76	4.68
SC3	C1	0.96	2.68	0.24	1.76
	C3	3.01	3.13	3.73	5.47
	C4	2.16	2.57	2.88	4.64
	C5	2.85	2.78	3.54	5.35
	C6	2.52	1.26	2.92	4.92
	C7	2.36	1.19	2.76	4.76
	Median	1.70	2.63	4.16	3.98

Note. The results are shown for sampling campaign six and three (SC6 and SC3), which are representative for early summer (i.e., after snowmelt) and typical summer conditions, respectively (see Figure 6). The results for the other campaigns were similar.

Areas with water Type II were mainly located at the upper hillslopes or ridge sites, or in the smaller catchments, and were characterized by high heavy metal concentrations and a relatively enriched isotopic composition. The more enriched composition of the groundwater again suggests that it is mainly recent precipitation. Concentrations of weathering-derived cations such as strontium and lithium were lower for some sites which could be the result of the shorter interaction times with the bedrock. The high heavy metal concentrations can be explained by two mechanisms. First, the soils on the forested ridges are more acidic than soils in the depressions as a result of intense weathering and decreased buffering capacity (Baltensweiler et al., 2017; Hagedorn et al., 2000; Schleppi et al., 1998). Because soil acidity and heavy metal adsorption are inversely related, more heavy metals are leached from the ridge soils (Harter, 1983), increasing the heavy metal concentrations in the ridge and hillslope groundwater. Second, heavy metals that have been introduced through atmospheric deposition (Shotyk, 1996; Shotyk, Blaser, Grünig, & Cheburkin, 2000) are likely to be more abundant on forested hillslopes and ridges due to dry deposition and evapo-concentration than at downslope locations, where they are more diluted by incoming rain-, soil-, and groundwater from further upslope. However, chloride concentrations did not show a similar evapo-concentration pattern to support the latter hypothesis.

The sites that had water Type III were located throughout the catchment and included most relatively deep wells. Some of the wells had a more depleted isotopic signature, and a smaller temporal variability in isotopic composition, others had above average EC values or calcium concentrations. The more depleted isotopic composition for some sites with this water type indicates that a larger proportion of recharge comes from snowmelt. The less variable isotopic composition suggests that there is more mixing or mixing within a larger reservoir for this groundwater so that the precipitation signal is more dampened. Therefore, we expect that groundwater in these wells is sustained by longer flowpaths. The higher weathering-derived cation concentrations (strontium and barium) at these sites also indicate longer interaction times with the bedrock material. Interestingly, the wells with a more depleted isotopic composition and higher strontium and barium concentrations did not have a higher EC, indicating that EC is in this case more strongly influenced by elements that are affected by different processes (as reflected by the high calcium and magnesium contributions to the ion balance).

The wells with water Type IV differed from the other wells by their high magnesium and sulfate concentrations, which were not related to any of the analysed landscape attributes. We hypothesize that these high concentrations are caused by local differences in the bedrock or as a result of geologic faults that bring a different part of the heterogeneous Flysch formation closer to the surface (Figure 1). One plausible explanation is that the composition of water Type IV is influenced by pyrite oxidation, which is known to increase sulfate concentrations and to reduce the pH (Brantley, Holleran, Jin, & Bazilevskaya, 2013). The enhanced acidity induces stronger dissolution of the carbonate bedrock and consequently increases the concentrations of base cations. A leaching experiment (c.f. Hissler et al., 2015) on a bedrock sample collected near streamgauge C1 (where streamflow was more similar to groundwater Type IV) and a sample collected near streamgauge C6 showed that the amount of magnesium that was released from the bedrock with increasing leachate acidity was much higher for the first bedrock sample than for the second. This could explain why magnesium concentrations were higher for this particular water type as well. A similar process might have caused the relatively high sulfate and magnesium concentrations in the streamwater samples compared with the groundwater (Figure 2c), as streamwater has a much higher oxidation capacity than ground water due to turbulent mixing and direct contact with the atmosphere.

Several previous studies have shown the usefulness of similar hydrogeomorphic units to describe run-off processes and catchment scale connectivity. Sidle et al. (2000) developed a hydrogeomorphic conceptual model for stormflow generation in the Hitachi Ohta Experimental Watershed in Japan based on hydrometric and tracer experiment data. They discretized the catchment into “hydrologically active” areas such as riparian zones, linear hillslope segments, and geomorphic hollows and showed that the contribution from each segment depends on the linkages (i.e., hydrologic connectivity) that exist during different stages of wetness. Similarly, McGlynn and McDonnell (2003) quantified the old water contributions from valley bottom riparian

zones and hillslopes for the Maimai catchment in New Zealand using hydrometric, isotopic, and solute data. With our study, we show that these units are also useful for describing catchment scale variability in groundwater chemistry during baseflow conditions. This knowledge can help to improve our understanding of catchment hydrologic connectivity and runoff generation processes and might help to identify source waters in catchments where no detailed spatial information on groundwater compositions is available.

4.3 | Which groundwater contributes to baseflow?

The considerable variability in the chemistry within the different groundwater types (Figure 6) complicates the calculations of the contributions from the different groundwater types to baseflow (Christophersen & Hooper, 1992; Pinder & Jones, 1969) and although methods to represent variability within end-members exist (Carrera, Vázquez-Suñé, Castillo, & Sánchez-Vila, 2004; Joerin, Beven, Iorgulescu, & Musy, 2002) would yield highly uncertain results. We therefore limited the analyses to qualitative interpretations, supported by the consistent differences in the element concentrations for the different groundwater types.

The comparison of the similarity between streamwater and the groundwater types (Table 6) suggests that during the wettest conditions (SC6, shortly after snowmelt), baseflow chemistry was more a mixture of the different groundwater types (Types I to III) than during drier conditions (SC3), when it resembled the riparian-like groundwater type (Type I) more. However, even during the wettest conditions, baseflow was not an equal mixture of the different groundwater types. During the majority of the campaigns, the streamwater chemistry most strongly resembled groundwater with a riparian-like chemistry, except at C1. The second highest similarity during most campaigns was to the "hillslope" groundwater (Type II). Streamflow was most similar to this hillslope groundwater during the campaigns just after snowmelt. We hypothesize that the wet conditions enable hillslope-riparian zone-stream connectivity (Jencso et al., 2009; von Freyberg et al., 2014), which leads to a larger similarity between hillslope groundwater and streamflow chemistry during these wetter conditions. This connectivity during wet conditions after snowmelt is also reflected in the stronger correlation between topography and water levels during this period (Rinderer et al., 2014).

The high spatial variability in streamwater chemistry suggests that the streams are not all fed by the same source areas. Our data suggest that during all campaigns streamflow at C1 was more strongly dominated by "deeper groundwater" (Type III) than the other sites. The location of a spring surrounded by exposed bedrock from recent landsliding is also an indication that upwelling of deep groundwater is an important process in this subcatchment. The similarity of streamwater to deeper groundwater (Type III) confirms the importance of deeper groundwater flow to sustain stream baseflow in conceptual models (Sklash, Farvolden, & Fritz, 1976; Wittenberg & Sivapalan, 1999) and source water studies (e.g., Mulholland, 1993; Tague, Grant, Farrell, Choate, & Jefferson, 2008). The closer similarity of streamwater to groundwater Types I and II suggests that the deeper groundwater component in the Studibach might be volumetrically less important, potentially due to the relatively "wet" characteristics of this catchment.

The resemblance of streamflow to the "riparian" groundwater (Type I) partly contradicts earlier results from baseflow sampling campaigns in the Alptal catchments (Fischer et al., 2015) that concluded that most streamflow was generated in the upper spring zone, an area with many springs that were assumed to be fed from "deeper groundwater." However, the isotope results suggest that these springs are fed by relatively shallow groundwater, as also reflected by their temperature and flow response to precipitation. These contrasting interpretations from different baseflow sampling campaigns could also be partly due to the different tracers used in the different studies (element concentrations and isotopes in this study vs. DOC, δ^2H, and calcium in Fischer et al., 2015) and highlight the overall sensitivity to the number of tracers used for the analyses (Barthold et al., 2011).

When interpreting these results, we should also not forget that the resemblance between groundwater and streamwater chemistry can only indicate which landscape features are potential sources of baseflow. Active contribution, which is assumed from hydrochemical similarity, needs to be confirmed with a measure of hydrological connectivity (Blume & van Meerveld, 2015). Additionally, similarity of streamwater chemistry to riparian-like groundwater could partly be the result of resetting of the chemical signature of the water for some of the solutes (e.g., sorption of the heavy metals from upslope water) before it enters the stream (Burt, 2005; Mulholland, 1992), which would lead to an overestimation of the riparian-like groundwater contribution to baseflow.

Because the resemblance of baseflow to the different groundwater types changed over time, the variability in groundwater chemistry should also be taken into account when determining the contributions of different end-members to streamflow. This becomes particularly clear when comparing the observed spatial variability in groundwater chemistry to the typical temporal variability in streamflow during events. The median spatial variability in the isotopic composition of the groundwater (spatial variability: median, 18.5‰ δ^2H; range, 11.9–23.7‰ δ^2H, respectively) was (much) larger than the typical change in stream water composition during rainfall events (median range in streamflow composition during six events in the Studibach catchment: 8.4‰ δ^2H; Fischer, Stähli, & Seibert, 2017, and median range for 24 events in the neighbouring Erlenbach catchment: 7.3‰ δ^2H; von Freyberg, Studer, Rinderer, & Kirchner, 2018). The expansion of the connected area towards the hillslopes (Zuecco et al., 2019) with wetter antecedent conditions will increase contributions from different storages, which will change the chemistry and isotopic composition of streamflow. As a result, one might underestimate or overestimate the amount of event water during stormflow when these other groundwater storages become hydrologically connected and contribute to streamflow, if the groundwater signature is only based on riparian-like groundwater type or a pre-event water sample that represents the riparian-like groundwater.

5 | CONCLUSION

So far, only a limited number of studies have quantified the spatial variability in shallow groundwater chemistry by directly sampling groundwater, trench outflow, or a combination of both at multiple locations. More often, groundwater is sampled at only a few locations or its chemical composition is inferred from baseflow, which is assumed to reflect the average groundwater in the catchment. We sampled shallow groundwater in 34–47 wells across the 20-ha Studibach catchment in Switzerland and found that the spatial variability in shallow groundwater chemistry is large and larger than the seasonal variation for most elements. These findings indicate that a few (riparian) groundwater samples or baseflow samples do not adequately characterize catchment scale shallow groundwater chemistry.

We could distinguish four shallow groundwater types and related these to topographic and hydrodynamic attributes. Groundwater Type I is characterized by high transition metal concentrations and is located mainly in the flat, wet locations with high groundwater tables. Groundwater Type II has high heavy metal concentrations and is located mainly on well drained, steeper slopes and in the smaller subcatchments. Groundwater Type III was found predominantly in relatively deep wells and is characterized by a more stable isotopic composition. Groundwater Type IV was found in two distinct areas of the catchment and is characterized by relatively high magnesium and sulfate concentrations. However, there was also considerable variability within each groundwater type. Baseflow was not an equal mixture of the different groundwater sources. During the majority of the campaigns, the baseflow chemistry most strongly resembled the riparian-like groundwater (Type I), except for a small subcatchment, where it resembled the deeper groundwater (Type III). The similarity to the hillslope-like groundwater (Type II) was larger during the campaigns after the snowmelt, suggesting increased connectivity and larger contributions of hillslope groundwater to streamflow during the wetter conditions. We expect that similar groundwater types can be found in other catchments with steep hillslopes and shallow groundwater levels and recommend sampling of all groundwater types to understand shallow groundwater chemistry and groundwater contributions to streamflow.

ACKNOWLEDGMENTS

This research project would not have been possible without the help and support of many people in the field and lab. We particularly thank Bjorn Studer and the Environmental Systems group at ETH Zürich for support with the chemical analyses; Barbara Herbstritt for the isotope analyses; Michael Rinderer for installing all the wells and the helpful discussions; and the Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Oberallmeindkorporation Schwyz (OAK), the municipality Alpthal, and the Department of Environment of the Canton of Schwyz for the good cooperation.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Leonie Kiewiet https://orcid.org/0000-0002-1437-1887
Jana von Freyberg https://orcid.org/0000-0002-2111-0001
H.J. (Ilja) van Meerveld https://orcid.org/0000-0002-7547-3270

REFERENCES

Anderson, M. G., & Burt, T. P. (1978). The role of topography in controlling throughflow generation. *Earth Surface Processes*, *3*, 331–344. https://doi.org/10.1002/esp.3290030402

Asano, Y., Uchida, T., Mimasu, Y., & Ohte, N. (2009). Spatial patterns of stream solute concentrations in a steep mountainous catchment with a homogeneous landscape. *Water Resources Research*, *45*(10), 1–9. https://doi.org/10.1029/2008WR007466

Baltensweiler, A., Walthert, L., Ginzler, C., Sutter, F., Purves, R. S., & Hanewinkel, M. (2017). Terrestrial laser scanning improves digital elevation models and topsoil pH modelling in regions with complex topography and dense vegetation. *Environmental Modelling and Software*, *95*, 13–21. https://doi.org/10.1016/j.envsoft.2017.05.009

Barthold, F. K., Tyralla, C., Schneider, K., Vaché, K. B., Frede, H. G., & Breuer, L. (2011). How many tracers do we need for end member mixing analysis (EMMA)? A sensitivity analysis. *Water Resources Research*, *47*(8), 1–14. https://doi.org/10.1029/2011WR010604

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, *24*(1), 43–69. https://doi.org/10.1080/02626667909491834

Blume, T., & van Meerveld, H. J. I. (2015). From hillslope to stream: Methods to investigate subsurface connectivity. *Wiley Interdisciplinary Reviews: Water*, *2*(3), 177–198. https://doi.org/10.1002/wat2.1071

Bracken, L. J., & Croke, J. (2007). The concept of hydrological connectivity and its contribution to understanding runoff-dominated geomorphic systems. *Hydrological Processes*, *21*, 2267–2274. https://doi.org/10.1002/hyp.6313

Brantley, S. L., Holleran, M. E., Jin, L., & Bazilevskaya, E. (2013). Probing deep weathering in the Shale Hills Critical Zone Observatory, Pennsylvania (USA): The hypothesis of nested chemical reaction fronts in the subsurface. *Earth Surface Processes and Landforms*, *38*(11), 1280–1298. https://doi.org/10.1002/esp.3415

Brinson M. M. (1993). A hydrogeomorphic classification for wetlands, wetlands research program technical report WRP-DE-4. Washington D.C.

Brown, V. A., McDonnell, J. J., Burns, D. A., & Kendall, C. (1999). The role of event water, a rapid shallow flow component, and catchment size in summer stormflow. *Journal of Hydrology*, *217*(3–4), 171–190. https://doi.org/10.1016/S0022-1694(98)00247-9

Burns, D. A., Hooper, R. P., McDonnell, J. J., Freer, J. E., Kendall, C., & Beven, K. (1998). Base cation concentrations in subsurface flow from a forested hillslope: The role of flushing frequency. *Water Resources Research*, *34*(12), 3535–3544. https://doi.org/10.1029/98WR02450

Burns, D. A., & McDonnell, J. J. (1998). Effects of a beaver pond on runoff processes: Comparison of two headwater catchments. *Journal of Hydrology*, *205*(3–4), 248–264. https://doi.org/10.1016/S0022-1694(98)00081-X

Burt, T. P. (2005). A third paradox in catchment hydrology and biogeochemistry: Decoupling in the riparian zone. *Hydrological Processes*, *19*(10), 2087–2089. https://doi.org/10.1002/hyp.5904

Buttle, M. (1994). Isotope hydrograph separations and rapid delivery of pre-event water from drainage basins. *Progress in Physical Geography*, *18*, 16–41. https://doi.org/10.1177/030913339401800102

Carrera, J., Vázquez-Suñé, E., Castillo, O., & Sánchez-Vila, X. (2004). A methodology to compute mixing ratios with uncertain end-members. *Water Resources Research*, *40*(12), 1–11. https://doi.org/10.1029/2003WR002263

Christophersen, N., & Hooper, R. P. (1992). Multivariate analysis of stream water chemical data: The use of principal components analysis for the end-member mixing problem. *Water Resources Research*, *28*(1), 99–107. https://doi.org/10.1029/91WR02518

Cirmo, C. P., & McDonnell, J. J. (1997). Linking the hydrologic and biogeochemical controls of nitrogen transport in near-stream zones of temperate-forested catchments: A review. *Journal of Hydrology*, *199*(1–2), 88–120. https://doi.org/10.1016/S0022-1694(96)03286-6

Detty, J. M., & McGuire, K. J. (2010). Topographic controls on shallow groundwater dynamics: Implications of hydrologic connectivity between hillslopes and riparian zones in a till mantled catchment. *Hydrological Processes*, *24*(16), 2222–2236. https://doi.org/10.1002/hyp.7656

Feyen, H., Wunderli, H., Wydler, H., & Papritz, A. (1999). A tracer experiment to study flow paths of water in a forest soil. *Journal of Hydrology*, *225*(3–4), 155–167. https://doi.org/10.1016/S0022-1694(99)00159-6

Fischer, B. M. C., Rinderer, M., Schneider, P., Ewen, T., & Seibert, J. (2015). Contributing sources to baseflow in pre-alpine headwaters using spatial snapshot sampling. *Hydrological Processes*, *29*(26), 5321–5336. https://doi.org/10.1002/hyp.10529

Fischer, B. M. C., Stähli, M., & Seibert, J. (2017). Pre-event water contributions to runoff events of different magnitude in pre-alpine headwaters. *Hydrology Research*, *48*(1), 28–47. https://doi.org/10.2166/nh.2016.176

Hagedorn, F., Schleppi, P., Waldner, P., & Flühler, H. (2000). Export of dissolved organic carbon and nitrogen from Gleysol dominated catchments—The significance of water flow paths. *Biogeochemistry*, *50*, 137–161. https://doi.org/10.1023/A:1006398105953

Harter, R. (1983). Effect of soil pH on adsorption of lead, copper, zinc, and nickel. *Soil Science Society of America Journal*, *47*, 47–51. https://doi.org/10.2136/sssaj1983.03615995004700010009x

Hill, A. R. (2000). Stream chemistry and riparian zones. In J. B. Jones, & P. J. Mulholland (Eds.), *Streams and ground waters* (pp. 83–110). San Diego, California: Academic Press.

Hissler, C., Stille, P., Juilleret, J., François, J., Perrone, T., & Morvan, G. (2015). Elucidating the formation of terra fuscas using Sr–Nd–Pb isotopes and rare earth elements. *Applied Geochemistry*, *54*, 85–99. https://doi.org/10.1016/j.apgeochem.2015.01.011

Hjerdt, K. N. (2004). A new topographic index to quantify downslope controls on local drainage. *Water Resources Research*, *40*(5), 1–6. https://doi.org/10.1029/2004WR003130

Hugenschmidt, C., Ingwersen, J., Sangchan, W., Sukvanachaikul, Y., Duffner, A., Uhlenbrook, S., & Streck, T. (2014). A three-component hydrograph separation based on geochemical tracers in a tropical mountainous headwater catchment in northern Thailand. *Hydrology and Earth System Sciences*, *18*(2), 525–537. https://doi.org/10.5194/hess-18-525-2014

Jencso, K. G., & McGlynn, B. L. (2011). Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation. *Water Resources Research*, *47*(11), 1–16. https://doi.org/10.1029/2011WR010666

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., & Marshall, L. A. (2009). Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale. *Water Resources Research*, *45*(4), 1–16. https://doi.org/10.1029/2008WR007225

Joerin, C., Beven, K. J., Iorgulescu, I., & Musy, A. (2002). Uncertainty in hydrograph separations based on geochemical mixing models. *Journal of Hydrology*, *255*(1–4), 90–106. https://doi.org/10.1016/S0022-1694(01)00509-1

Kendall, C., & Mcdonnell, J. J. (1998). *Isotope tracers in catchment hydrology*. Amsterdam: Elsevier Science Publishers.

Kendall, C., McDonnell, J. J., & Gu, W. (2001). A look inside 'black box' hydrograph separation models: A study at the hydrohill catchment. *Hydrological Processes*, *15*(10), 1877–1902. https://doi.org/10.1002/hyp.245

Kirchner, J. W. (2003). A double paradox in catchment hydrology and geochemistry. *Hydrological Processes*, *17*(4), 871–874. https://doi.org/10.1002/hyp.5108

Kirchner, J. W., & Neal, C. (2013). Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(30), 12213–12218. https://doi.org/10.1073/pnas.1304328110

Klaus, J., & McDonnell, J. J. (2013). Hydrograph separation using stable isotopes: Review and evaluation. *Journal of Hydrology*, *505*, 47–64. https://doi.org/10.1016/j.jhydrol.2013.09.006

Laudon, H., & Slaymaker, O. (1997). Hydrograph separation using stable isotopes, silica and electrical conductivity: An alpine example. *Journal of Hydrology*, *201*, 82–101.

Lidman, F., Boily, Å., Laudon, H., & Köhler, S. J. (2017). From soil water to surface water—How the riparian zone controls element transport from a boreal forest to a stream. *Biogeosciences*, *14*, 3001–3014. https://doi.org/10.5194/bg-14-3001-2017

Lindsay, J. B. (2016). Whitebox GAT: A case study in geomorphometric analysis. *Computers and Geosciences*, *95*, 75–84. https://doi.org/10.1016/j.cageo.2016.07.003

Loheide, S. P., & Gorelick, S. M. (2006). Quantifying stream-aquifer interactions through the analysis of remotely sensed thermographic profiles and in situ temperature histories. *Environmental Science and Technology*, *40*(10), 3336–3341. https://doi.org/10.1021/es0522074

McGlynn, B., McDonnell, J., Stewart, M., & Seibert, J. (2003). On the relationships between catchment scale and streamwater mean residence time. *Hydrological Processes*, *17*(1), 175–181. https://doi.org/10.1002/hyp.5085

McGlynn, B. L., & McDonnell, J. J. (2003). Quantifying the relative contributions of riparian and hillslope zones to catchment runoff. *Water Resources Research*, *39*(11). https://doi.org/10.1029/2003WR002091

McKenzie, R. M. (1980). The adsorption of lead and other heavy metals on oxides of manganese and iron. *Australian Journal of Soil Research 18*(1) 61–73. https://doi.org/10.1071/SR9800061

Megonigal, J. P., Patrick, W. H., & Faulkner, S. P. (1993). Wetland identification in seasonally flooded forest soils: Soil morphology and redox dynamics. *Soil Science Society of America Journal*, *57*(1), 140. https://doi.org/10.2136/sssaj1993.03615995005700010027x

Mohn, J., Schürmann, A., Hagedorn, F., Schleppi, P., & Bachofen, R. (2000). Increased rates of denitrification in nitrogen-treated forest soils. *Forest Ecology and Management*, *137*(1–3), 113–119. https://doi.org/10.1016/S0378-1127(99)00320-5

Molnar, P., Densmore, A. L., McArdell, B. W., Turowski, J. M., & Burlando, P. (2010). Analysis of changes in the step-pool morphology and channel profile of a steep mountain stream following a large flood. *Geomorphology*, *124*(1–2), 85–94. https://doi.org/10.1016/j.geomorph.2010.08.014

Mulholland, P. J. (1992). Regulation of nutrient concentrations in a temperate forest stream: Roles of upland, riparian, and instream Processes. *Limnology and Oceanography*, *37*(7), 1512–1526. https://doi.org/10.4319/lo.1992.37.7.1512

Mulholland, P. J. (1993). Hydrometric and stream chemistry evidence of three storm flowpaths in Walker Branch Watershed. *Journal of*

Hydrology, *151*(2–4), 291–316. https://doi.org/10.1016/0022-1694(93)90240-A

Ocampo, C. J., Sivapalan, M., & Oldham, C. (2006). Hydrological connectivity of upland-riparian zones in agricultural catchments: Implications for runoff generation and nitrate transport. *Journal of Hydrology*, *331*(3–4), 643–658. https://doi.org/10.1016/j.jhydrol.2006.06.010

Ogawa, A., Shibata, H., Suzuki, K., Mitchell, M. J., & Ikegami, Y. (2006). Relationship of topography to surface water chemistry with particular focus on nitrogen and organic carbon solutes within a forested watershed in Hokkaido, Japan. *Hydrological Processes*, *20*(2), 251–265. https://doi.org/10.1002/hyp.5901

Penna, D., Mantese, N., Hopp, L., Dalla Fontana, G., & Borga, M. (2015). Spatio-temporal variability of piezometric response on two steep alpine hillslopes. *Hydrological Processes*, *29*(2), 198–211. https://doi.org/10.1002/hyp.10140

Penna, D., van Meerveld, H. J., Zuecco, G., Dalla Fontana, G., & Borga, M. (2016). Hydrological response of an Alpine catchment to rainfall and snowmelt events. *Journal of Hydrology*, *537*, 382–397. https://doi.org/10.1016/j.jhydrol.2016.03.040

Peralta-Tapia, A., Sponseller, R. A., Ågren, A., Tetzlaff, D., Soulsby, C., & Laudon, H. (2015). Scale-dependent groundwater contributions influence patterns of winter baseflow stream chemistry in boreal catchments. *Journal of Geophysical Research: Biogeosciences*, *120*, 847–858. https://doi.org/10.1002/2014JG002878

Pinder, G. F., & Jones, J. F. (1969). Determination of the ground-water component of peak discharge from the chemistry of total runoff. *Water Resources Research*, *5*(2), 438–445. https://doi.org/10.1029/WR005i002p00438

Pint, C. D., Hunt, R. J., & Anderson, M. P. (2003). Flowpath delineation and groud water age Allquash Basin, Wisconsin. *Ground Water*, *41*(7), 895–902. https://doi.org/10.1111/j.1745-6584.2003.tb02432.x

Rinderer, M., McGlynn, B. L., & van Meerveld, H. J. (2017). Groundwater similarity across a watershed derived from time-warped and flow-corrected time series. *Water Resources Research*, *53*, 3921–3940. https://doi.org/10.1002/2016WR019856

Rinderer, M., van Meerveld, H. J., & Seibert, J. (2014). Topographic controls on shallow groundwater levels in a steep, prealpine catchment: When are the TWI assumptions valid? *Water Resources Research*, *50*, 6067–6080. https://doi.org/10.1002/2013WR015009

Rinderer, M., van Meerveld, I., Stähli, M., & Seibert, J. (2015). Is groundwater response timing in a pre-alpine catchment controlled more by topography or by rainfall? *Hydrological Processes*, *30*(7), 1036–1051. https://doi.org/10.1002/hyp.10634

Schleppi, P., Muller, N., Feyen, H., Papritz, A., Bucher, J. B., & Flühler, H. (1998). Nitrogen budgets of two small experimental forested catchments at Alptal, Switzerland. *Forest Ecology and Management*, *101*(1–3), 177–185. https://doi.org/10.1016/S0378-1127(97)00134-5

Schleppi, P., Waldner, P. A., & Fritschi, B. (2006). Accuracy and precision of different sampling strategies and flux integration methods for runoff water: Comparisons based on measurements of the electrical conductivity. *Hydrological Processes*, *20*(2), 395–410. https://doi.org/10.1002/hyp.6057

Seibert, J., & McGlynn, B. L. (2007). A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. *Water Resources Research*, *43*(4), 1–8. https://doi.org/10.1029/2006WR005128

Shotyk, W. (1996). Natural and anthropogenic enrichments of As, Cu, Pb, Sb, and Zn in ombrotrophic versus minerotrophic peat bog profiles, Jura Mountains, Switzerland. *Water, Air, and Soil Pollution*, *90*(3–4), 375–405. https://doi.org/10.1007/BF00282657

Shotyk, W., Blaser, P., Grünig, A., & Cheburkin, A. K. (2000). A new approach for quantifying cumulative, anthropogenic, atmospheric lead deposition using peat cores from bogs: Pb in eight Swiss peat bog profiles. *Science of the Total Environment*, *249*(1–3), 281–295. https://doi.org/10.1016/S0048-9697(99)00523-9

Sidle, R. C., Tsuboyama, Y., Noguchi, S., Hosoda, I., Fujieda, M., & Shimizu, T. (2000). Stormflow generation in steep forested headwaters: A linked hydrogeomorphic paradigm. *Hydrological Processes*, *14*, 369–385. https://doi.org/10.1002/(SICI)1099-1085(20000228)14:3<369::AID-HYP943>3.3.CO;2-G

Sklash, M. G., Farvolden, R. N., & Fritz, P. (1976). A conceptual model of watershed response to rainfall, developed through the use of oxygen-18 as a natural tracer. *Canadian Journal of Earth Sciences*, *13*(2), 271–283. https://doi.org/10.1139/e76-029

Soulsby, C., Chen, M., Ferrier, R. C., Helliwell, R. C., Jenkins, A., & Harriman, R. (1998). Hydrogeochemistry of shallow groundwater in an upland Scottish catchment. *Hydrological Processes*, *12*(7), 1111–1127. https://doi.org/10.1002/(SICI)1099-1085(19980615)12:7<1111::AID-HYP633>3.0.CO;2-2

Soulsby, C., Tetzlaff, D., Rodgers, P., Dunn, S., & Waldron, S. (2006). Runoff processes, stream water residence times and controlling landscape characteristics in a mesoscale catchment: An initial evaluation. *Journal of Hydrology*, *325*(1–4), 197–221. https://doi.org/10.1016/j.jhydrol.2005.10.024

Stähli, M., & Gustafsson, D. (2006). Long-term investigations of the snow cover in a subalpine semi-forested catchment. *Hydrological Processes*, *20*(2), 411–428. https://doi.org/10.1002/hyp.6058

Stutter, M. I., Deeks, L. K., Low, D., & Billett, M. F. (2006). Impact of soil and groundwater heterogeneity on surface water chemistry in an upland catchment. *Journal of Hydrology*, *318*(1–4), 103–120. https://doi.org/10.1016/j.jhydrol.2005.06.007

Tague, C., Grant, G., Farrell, M., Choate, J., & Jefferson, A. (2008). Deep groundwater mediates streamflow response to climate warming in the Oregon Cascades. *Climatic Change*, *86*(1–2), 189–210. https://doi.org/10.1007/s10584-007-9294-8

Tetzlaff, D., Soulsby, C., Waldron, S., Malcolm, I. A., Bacon, P. J., Dunn, S. M., … Youngson, A. F. (2007). Conceptualization of runoff processes using a geographical information system and tracers in a nested mesoscale catchment. *Hydrological Processes*, *21*(10), 1289–1307. https://doi.org/10.1002/hyp.6309

van Meerveld, H. J. I., Fischer, B. M. C., Rinderer, M., Stähli, M., & Seibert, J. (2018). Runoff generation in a pre-alpine catchment: A discussion between a tracer and a shallow groundwater hydrologist. *Cuadernos de Investigación Geográfica*, *44*(2), 429–452. https://doi.org/10.18172/cig.3349

von Freyberg, J., Radny, D., Gall, H. E., & Schirmer, M. (2014). Implications of hydrologic connectivity between hillslopes and riparian zones on streamflow composition. *Journal of Contaminant Hydrology*, *169*, 62–74. https://doi.org/10.1016/j.jconhyd.2014.07.005

von Freyberg, J., Studer, B., Rinderer, M., & Kirchner, J. W. (2018). Studying catchment storm response using event- and pre-event-water volumes as fractions of precipitation rather than discharge. *Hydrology and Earth System Sciences*, *22*(11), 5847–5865. https://doi.org/10.5194/hess-22-5847-2018

Walker, J. F., Hunt, R. J., Bullen, T. D., Krabbenhoft, D. P., & Kendall, C. (2003). Variability of isotope and major ion chemistry in the Allequash Basin, Wisconsin. *GroundWater*, *41*(7), 883–894. https://doi.org/10.1111/j.1745-6584.2003.tb02431.x

Weiler, M., & McDonnell, J. J. (2006). Testing nutrient flushing hypotheses at the hillslope scale: A virtual experiment approach. *Journal of Hydrology*, *319*(1–4), 339–356. https://doi.org/10.1016/j.jhydrol.2005.06.040

Welsch, D. L., Kroll, C. N., McDonnell, J. J., & Burns, D. A. (2001). Topographic controls on the chemistry of subsurface stormflow. *Hydrological Processes*, *15*(10), 1925–1938. https://doi.org/10.1002/hyp.247

Wittenberg, H., & Sivapalan, M. (1999). Watershed groundwater balance estimation using streamflow recession analysis and baseflow

Appendix

separation. *Journal of Hydrology*, *219*(1–2), 20–33. https://doi.org/10.1016/S0022-1694(99)00040-2

Zuecco, G., Rinderer, M., Penna, D., Borga, M., & van Meerveld, H. J. (2019). Quantification of subsurface hydrologic connectivity in four headwater catchments using graph theory. *Science of the Total Environment*, *646*, 1265–1280. https://doi.org/10.1016/j.scitotenv.2018.07.269

How to cite this article: Kiewiet L, von Freyberg J, van Meerveld H.J. (Ilja). Spatiotemporal variability in hydrochemistry of shallow groundwater in a small pre-alpine catchment: The importance of landscape elements. *Hydrological Processes*. 2019;1–21. https://doi.org/10.1002/hyp.13517

B: Paper II

Hydrol. Earth Syst. Sci., 24, 3381–3398, 2020
https://doi.org/10.5194/hess-24-3381-2020
© Author(s) 2020. This work is distributed under
the Creative Commons Attribution 4.0 License.

Hydrology and Earth System Sciences
Open Access
EGU

Do stream water solute concentrations reflect when connectivity occurs in a small, pre-Alpine headwater catchment?

Leonie Kiewiet[1], **Ilja van Meerveld**[1], **Manfred Stähli**[2], **and Jan Seibert**[1,3]

[1]Department of Geography, University of Zurich, Zurich, Switzerland
[2]Swiss Federal Research Institute WSL, Birmensdorf, Switzerland
[3]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden

Correspondence: Leonie Kiewiet (leonie.kiewiet@geo.uzh.ch)

Received: 28 December 2019 – Discussion started: 14 January 2020
Revised: 7 May 2020 – Accepted: 29 May 2020 – Published: 2 July 2020

Abstract. Expansion of the hydrologically connected area during rainfall events causes previously disconnected areas to contribute to streamflow. If these newly contributing areas have a different hydrochemical composition compared to the previously connected contributing areas, this may cause a change in stream water chemistry that cannot be explained by simple mixing of rainfall and baseflow. Changes in stormflow composition are, therefore, sometimes used to identify when transiently connected areas (or water sources) contribute to stormflow. We identified the dominant sources of streamflow for a steep 20 ha pre-Alpine headwater catchment in Switzerland and investigated the temporal changes in connectivity for four rainfall events based on stream water concentrations and groundwater level data. First, we compared the isotopic and chemical composition of stormflow at the catchment outlet to the composition of rainfall, groundwater and soil water. Three-component end-member mixing analyses indicated that groundwater dominated stormflow during all events, and that soil water fractions were minimal for three of the four events. However, the large variability in soil and groundwater composition compared to the temporal changes in stormflow composition inhibited the determination of the contributions from the different groundwater sources. Second, we estimated the concentrations of different solutes in stormflow based on the mixing fractions derived from two-component hydrograph separation using a conservative tracer (δ^2H) and the measured concentrations of the solutes in baseflow and rainfall. The estimated concentrations differed from the measured stormflow concentrations for many solutes and samples. The deviations increased gradually with increasing streamflow for some solutes (e.g. iron and copper), suggesting increased contributions from riparian and hillslope groundwater with higher concentrations of these solutes and thus increased hydrological connectivity. The findings of this study show that solute concentrations partly reflect the gradual changes in hydrologic connectivity, and that it is important to quantify the variability in the composition of different source areas.

1 Introduction

During dry periods only a small part of a catchment is connected to the stream, but the connected area can expand dramatically during rainfall or snowmelt events (Stieglitz et al., 2003; Bracken and Croke, 2007; Jencso and McGlynn, 2011; van Meerveld et al., 2015). Knowledge of which areas are connected and contribute to streamflow is important because it helps us to shape our conceptual understanding of how catchments function. For example, Ladouche et al. (2001) showed for the 0.8 km^2 Strengbach catchment in France that the upper layers of saturated areas contributed up to 30 % of the discharge during the initial stages of a rainfall event, even though these areas occupied only 2 % of the catchment area. However, during the final stage of the event, upslope and downslope areas contributed equally to flow. Similarly, Oswald et al. (2011), showed for a 0.8 km^2 catchment in northwestern Ontario, Canada, that a large part of the catchment area was hydrologically disconnected from the stream during most events, and that there was a threshold catchment storage at which a larger area contributed to streamflow. Connection of upslope areas does not only lead to large changes in

Published by Copernicus Publications on behalf of the European Geosciences Union.

discharge (Lehmann et al., 2007; Detty and McGuire, 2010; van Meerveld et al., 2015) but can also cause major changes in stream water composition (e.g. Devito and Hill, 1997; Stieglitz et al., 2003; Ocampo et al., 2006). Interpretations of hydrologic connectivity are often based on such changes in stream water chemistry (Uhlenbrook et al., 2004; Soulsby et al., 2007; Pacific et al., 2010).

Hydrologic connectivity, i.e. "the linkage of separate regions of a catchment via water flow" (Blume and van Meerveld, 2015), is usually inferred from either stream-based or hillslope-based measurements because direct observations of connectivity are limited due to the difficulty in observing and quantifying subsurface processes (Hopp and McDonnell, 2009; Blume and van Meerveld, 2015). In many studies, conservative tracers (e.g. stable water isotopes or non-reactive elements) are selected to identify the origin of streamflow, using methods such as hydrograph separation (Buttle, 1994) or end-member mixing analyses (EMMA; Hooper et al., 1990; Christophersen and Hooper, 1992). Tracers can also be used to assess connectivity of hillslopes to the streams (Tetzlaff et al., 2014; Uhlenbrook et al., 2004). Since stream chemistry is the proportional mixture of all actively contributing areas, quantifying each contribution results in a measure for catchment-wide connectivity. For instance, McGlynn and McDonnell (2003) used silica concentrations and isotope data for a 2.6 ha sub-catchment of the Maimai catchment in New Zealand to show that the contributions from the hillslopes were larger for an event with higher wetness conditions than for an event with drier initial conditions and were also larger on the falling limb of the hydrograph. Several studies in the 31 km^2 Girnock Burn catchment in Scotland investigated the connectivity of source areas to the stream using Gran alkalinity and isotope data (e.g. Soulsby et al., 2007; Tetzlaff et al., 2014). They found that the upper soil layers and upslope areas increasingly dominated streamflow at higher flows, and that the riparian peat soils modulated the stream water isotopic composition. However, few studies have compared the results from stream-based and hillslope-based inferences of connectivity. Burns et al. (1998) showed that hillslope contributions to streamflow inferred from end-member mixing analyses were similar to the subsurface flow measurements for a trenched hillslope.

Mixing analyses are traditionally performed with conservative solutes and stable water isotopes (Hooper and Shoemaker, 1986). Non-conservative solute concentrations can also provide useful information on hydrological connectivity and flow pathways because they can aid the identification of different source areas (Barthold et al., 2011; Abbott et al., 2018). The concentrations of specific elements can also be indicative for differences in redox conditions (e.g. sulfate, iron and manganese), bedrock-contact time (e.g. calcium, magnesium, sodium and barium), or vegetation (e.g. nitrogen, phosphorus and potassium; Kaushal et al., 2018). It has been suggested that the discrepancy between hydrograph separation results for conservative and non-conservative tracers highlights when and where stream water is not the result of conservative mixing between end members, such as baseflow and precipitation (Kirchner, 2003). Instead, it might reflect mixing from different "old" water sources in the catchment that have different concentrations. Therefore, this discrepancy may provide information on when hillslope–stream connectivity is established. Alternatively, the differences in the relative response of conservative and non-conservative tracers during rainfall events might be (partly) due to reactive processes that mobilise (or immobilise) solutes at the event timescale (Godsey et al., 2009). As such, focusing on solute responses in stormflow and the difference between conservative and non-conservative tracers might allow us to identify the extent of these reactive transport processes and contributions from "old" water sources that do not contribute to baseflow.

Solute concentrations in stream water might be relatively constant (chemostatic), decrease (dilution) or increase (mobilisation) in response to rainfall, depending on the source areas to streamflow and their respective concentrations, as well as reactive transport processes (Godsey et al., 2009; Seibert et al., 2009; Knapp et al., 2020). Godsey et al. (2009) found that concentrations of typical weathering products (calcium, magnesium, silica and sodium) were nearly chemostatic for 59 geochemically diverse US catchments, suggesting a (constant) source of these solutes. This implies that the areas that contribute to streamflow during rainfall events have similar concentrations of these solutes as the permanently contributing areas, higher concentrations to compensate for the dilution caused by the rainfall, or that reactions are fast enough to maintain similar concentrations during the event.

The timing of the onset of contributions from different source areas also affects the solute concentrations (Abbott et al., 2018). Several studies have shown that the relationship between concentrations and discharge is hysteretic at the event timescale (e.g. Evans and Davies, 1998; Hornberger et al., 2001). Zuecco et al. (2019) showed that the increase in subsurface connectivity was delayed compared to streamflow (anticlockwise hysteresis) for two sub-catchments of the Studibach catchment in Switzerland, suggesting that hillslope runoff may not be the dominant runoff source at the beginning of rainfall events for these small catchments. If hillslope and riparian zone water have a different composition, this can cause hysteresis in the relationship between solute concentrations and streamflow. Changes in solute concentrations might also depend on the size of the catchment (Brown et al., 1999) and mixing that occurs during transport from the source areas to the outlet. For instance, hillslope runoff may bypass the riparian zone through focused locations along the stream channel or via preferential flow pathways (Allaire et al., 2015) and mix with other hillslope sources (Seibert et al., 2009) and riparian groundwater (McGlynn and McDonnell, 2003; Chanat and Hornberger, 2003) on its way to the stream.

For all analyses of source areas and connectivity, it is important to quantify the variability in the concentrations of conservative and non-conservative tracers because it affects the robustness of the results and thus interpretations of connectivity. However, for most small ($< 10\,km^2$) catchment studies it remains unclear how large the changes in stream water composition are compared to the spatial variability in groundwater and soil water because the spatial variability in groundwater and soil water is rarely assessed (Penna and van Meerveld, 2019). In this study, we combined spatially distributed soil- and groundwater sampling with event-based stream water sampling in the pre-Alpine Studibach catchment to address the following research questions:

1. How variable is stream water chemistry during events compared to the spatial variability in soil and groundwater chemistry?
2. What are the dominant sources of streamflow during small to intermediately sized rainfall events?
3. How much do the changes in the concentrations of conservative and non-conservative tracers differ during events, and does this difference provide information on the relative contributions of different parts of the catchment and, thus, hydrological connectivity?

2 Study catchment

We conducted this study in the $0.2\,km^2$ pre-Alpine Studibach catchment, a headwater catchment of the Zwäckentobel, located in the Alptal, canton Schwyz, Switzerland. The elevation of the Studibach ranges from 1270 to 1650 m above sea level. The mean annual precipitation is about $2300\,mm\,yr^{-1}$. The precipitation is relatively evenly distributed throughout the year (Feyen et al., 1999), and about one-third falls as snow (Stähli and Gustafsson, 2006). The catchment is steep (average slope: 35°) and characterised by a step-wise topography, with flatter areas and steep slopes due to soil creep and landslides. An open coniferous forest covers about half of the catchment (Hagedorn et al., 2000), a third is characterised as a moor landscape or wet grassland, and the remaining areas are alpine meadows.

Streamflow and groundwater levels respond quickly to rainfall (Fischer et al., 2015; Rinderer et al., 2015). The groundwater level response time is generally less than 30 min (Rinderer et al., 2014), and only 3 mm of cumulative rainfall already causes an increase in the groundwater level for a large part of the catchment during typical conditions (Rinderer et al., 2015). The groundwater level peak precedes the peak discharge in the Studibach at half of the sites but only by 15 or 20 min (Rinderer et al., 2015). Water levels in flatter locations and topographic depressions rise nearly instantaneously, which suggests that they can contribute to streamflow during the early stages of a rainfall event. Previous studies suggest that event water fractions in stormflow are generally low (Kiewiet et al., 2020; von Freyberg et al., 2018), except for events with more than 50 mm of rainfall (Fischer et al., 2017).

Soils are generally shallow (0.5 m at ridge sites to ~ 2.5 m in depressions); soil depth is weakly correlated to slope (van Meerveld et al., 2018). The gleysols are underlain by three different types of Flysch bedrock, which is a reworked carbonate rock consisting of deep-water deposits. The carbonate-rich bedrock results in high groundwater concentrations with a calcium-bicarbonate signature, although some sites have high sulfate and magnesium concentrations (Kiewiet et al., 2019).

The Studibach can be subdivided into four different landscape elements with a distinct groundwater composition (Kiewiet et al., 2019 and Fig. 1):

1. Riparian zone, flatter areas and topographic hollows with above-average concentrations of iron and manganese. These areas are from here on referred to as "riparian".
2. Hillslopes and steeper areas, characterised by above-average concentrations of copper, zinc and lead.
3. Areas with above-average concentrations of weathering-derived solutes such as strontium, indicative of longer (and deeper) flow pathways, which are from here on referred to as deep groundwater.
4. Areas located in a specific part of the catchment that is characterised by high magnesium and sulfate concentrations.

3 Methods

3.1 Hydrometric measurements

To monitor stream water and groundwater levels, we used a network of 51 shallow groundwater wells and streamflow gauges (Fig. 1) that was installed in 2009–2010 (Rinderer et al., 2014). The wells were distributed based on the topographic wetness index (TWI; Beven and Kirkby, 1979) and cover the range of wet and dry locations in the catchment. All wells were drilled by hand to the bedrock (0.5 to 2.5 m depth), screened over the entire length, except for the top 10 cm, and sealed with a layer of bentonite clay. Stream stage was measured directly in the stream (outlet; Fig. 1a) or behind a V-notch weir (C5). Water levels were measured at each well and stream location with either a capacitance water level logger (Odyssey Dataflow Systems Limited) or a pressure transducer (DCX-22 CTD Keller AG für Druckmesstechnick or STS DL/N 70, Sensor Technik Sirnach AG). The pressure data were corrected for changes in barometric pressure and temperature using the data from the MeteoSwiss station in Einsiedeln (910 m a.s.l; ca. 10 km from the catchment outlet). Rainfall was recorded at three locations within the catchment

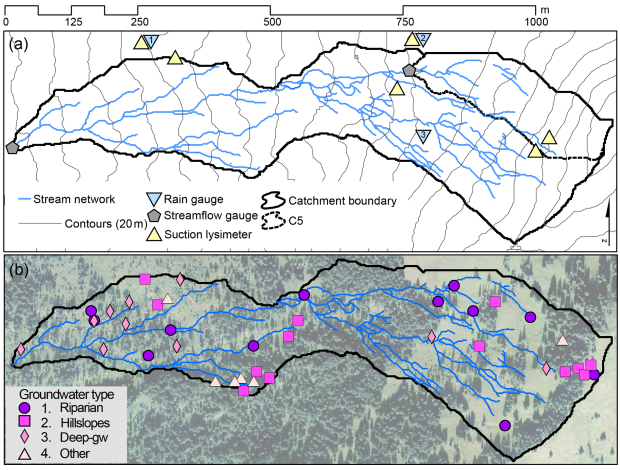

Figure 1. Maps of the Studibach catchment with **(a)** the stream network (blue lines), stream gauges (grey pentagons), rain gauges (blue triangles; 1–3) and suction lysimeters (yellow triangles), 20 m contour lines (grey), and the boundary of the catchment (black). C5 subcatchment (broken lines) and **(b)** location of the wells, colour-coded by groundwater type as follows: (1) riparian wells; (2) hillslope wells; (3) "deep" groundwater wells; and (4) wells with high magnesium and sulfate concentrations (based on Kiewiet et al., 2019).

with tipping-bucket rain gauges (0.2 mm resolution; Odyssey Dataflow Systems Limited; Fig. 1a).

The stream stage data were converted to specific discharge (Q, further referred to as discharge) using a rating curve based on 20 salt dilution measurements. Due to technical issues, there were no observations of stage height at the catchment outlet during events I and II (see Sect. 3.2). We used the correlation between the specific discharge at the catchment outlet and an intermediately sized sub-catchment (C5; Fig. 1a) for the four months following events I and II to estimate the streamflow at the outlet for the period without data (coefficient of determination $r^2 = 0.66$; RMSE = $0.75\,\mathrm{mm\,h^{-1}}$; for comparison, the 10th and 90th percentile of Q at the catchment outlet for this period were 0.35 and $2.11\,\mathrm{mm\,h^{-1}}$, respectively). We assume that the uncertainty in the discharge for events I and II does not affect our conclusions as they are largely based on relative changes in discharge during the events. The ranking of the events based on the peak of the (reconstructed) discharge was the same as the ranking based on the peak rainfall intensity.

3.2 Sample collection

We analysed streamflow and stream chemistry for four events (I–IV; Table 1) in the fall seasons of 2016 and 2017. Stream water samples were collected at the outlet of the Studibach using automatic samplers (full-size portable sampler, model no. 3712, Teledyne ISCO, USA). The sampling interval was based on the expected event duration. The multi-interval programme was set to sample stream water every 10 to 20 min at the start of the rising limb (maximum of six samples). The remaining 18 samples were taken at an hourly interval. We emptied the samplers within 24 h after sample collection to avoid fractionation. We used a timer to start the sampler if the expected time of the onset of the rainfall was during the night. Rainfall was collected with passive sequential samplers (built after Kennedy et al., 1979, and described in detail in Fischer et al., 2019) at two locations in the catchment (rain gauge location one and two in Fig. 1a). The samplers collected a sample for approximately every 5 mm of rainfall.

For soil water and groundwater, we used the data from a subset of nine baseflow snapshot campaigns during the snow-free seasons of 2016 and 2017 (Kiewiet et al., 2019). Soil water was collected with six to 18 suction lysimeters at four to six sites (at 15, 30 and 50 cm below the surface at forested and non-forested sites at three different elevations, namely 1361, 1502, 1611 m a.s.l.; Fig. 1a). We applied a tension of 50 mbar to the lysimeters and collected the soil water sample the next day. Groundwater was collected at all wells that contained water (34 to 38 wells). The shallow wells were either purged or at least twice the well volume was extracted a day before the sampling. For a detailed description of the groundwater sampling procedure, see Kiewiet et al. (2019).

Table 1. Overview of the four events analysed in this study: event duration (D, h), rainfall amount (P, mm), average and maximum 10 min rainfall intensity (I_p and $I_{p\text{-}max}$, mm h^{-1}), the maximum change in specific discharge (ΔQ, mm h^{-1}), the maximum change in isotopic composition of the stream water (δ^2H, ‰), the minimum and maximum fraction of the catchment that was connected (A_{min}-A_{max}) during the event, and the date of the groundwater and soil water sampling campaign.

Event	Start date	D (h)	P (mm)	I_p (mm h^{-1})	$I_{p\text{-}max}$ (mm h^{-1})	ΔQ (mm h^{-1})	$Q-\delta^2$H (‰)	A_{min}-A_{max} (–)	Date of sampling campaign
I	2 October 2016	14	17	1.2	7	0.02–0.07	−70.5 to −65.7	0.27–0.48	5 October 2016
II	25 October 2016	28	33	1.2	13	0.02–0.17	−75.3 to −67.6	0.27–0.35*	5 October 2016
III	3 October 2017	7	27	3.9	24	0.08–0.43	−73.7 to −69.1	0.33–0.68	12 October 2017
IV	5 October 2017	27	32	1.2	10	0.07–0.30	−69.1 to −65.2	0.33–0.67	12 October 2017

* The fraction of the catchment that was hydrologically connected increased from 0.27 to 0.28 during the sampling period and to 0.35 during a discharge peak that occurred after the samplers stopped (see Fig. S3).

3.3 Sample analyses

The samples for cation and anion analyses were stored in a fridge (6 °C) before lab analyses (within a few days) or were frozen (−18 °C) directly after collection until shortly before the analyses. The samples were filtered (0.45 µm; SimplepureTM syringe filter) and acidified (only for cation analysis) to mobilise trace metals. The samples were analysed at the Physics of Environmental Systems laboratory at ETH Zurich (Switzerland) using an ion chromatograph (861 Advanced Compact IC, Metrohm AG, Switzerland) for anions and a mass spectrometer (7900 ICP-MS, Agilent, USA) for cations. Calibration curves were obtained from measurements with five calibration standards before or after measuring the samples.

The samples were analysed for stable water isotope composition with a cavity ring-down spectroscope (CRDS; L2140-i or L2130-i, Picarro, Inc., USA) at the Chair of Hydrology at the University of Freiburg (Germany). The reported precision is ±0.16 ‰ for δ^{18}O and ±0.6 ‰ for δ^2H. All samples plotted close to the local meteoric water line. The average (± standard deviation) of the line-conditioned excess (lc excess; Landwehr and Coplen, 2006) for all 516 stream, soil and groundwater samples was 5.3 ± 1.3 ‰, excluding five soil water samples (taken at 15 cm – three samples; 30 cm – one sample; and 50 cm – one sample, respectively, below the soil surface) for which lc excess ranged from −9.6 to −1.5 ‰. Deuterium excess (D_{ex}) was calculated as $D_{ex} = \delta^2H - (8 \cdot \delta^{18}O)$.

3.4 Data analysis

3.4.1 Relative concentrations

We examined the changes in stream water concentrations during the rainfall events using concentration–discharge (C–Q) relationships and identified the corresponding hysteresis index (cf. Zuecco et al., 2016). For this, we normalised both the discharge and the concentrations so that 0 represents the smallest measured value, and 1 the highest measured value.

For each solute, we calculated the relative concentration R_x by comparing the concentration of the sample to that of baseflow as follows:

$$R_x = \frac{C_{Q\_x}}{C_{BF\_x}}, \quad (1)$$

where C_{Q_x} and C_{BF_x} are the concentration of solute x in stream water during the event and in baseflow before the event, respectively. We define baseflow as the streamflow between rainfall-runoff events and assume that it comes from groundwater. The relative concentration indicates dilution ($R_x < 1$) or enrichment ($R_x \geq 1$) during the events. It thus quantifies the direction and magnitude of the change in solute concentrations (note that R_x is not an alternative measure for the fraction of baseflow in stormflow). We used the relative concentrations (R_x; Eq. 2) to identify groups of solutes using hierarchical cluster analysis.

3.4.2 Hydrograph separation and end-member mixing analysis

We tested if the median concentrations of different (ground-)water types were significantly different (Table 2; Tukey–Kramer test; Tukey HSD test in the "agricolae" R package). We pairwise tested seven groups, namely all groundwater, riparian groundwater, hillslope groundwater, all soil water, soil water at forested sites, soil water at non-forested sites and rainfall. We performed all computations in R (R Core Team, 2013) and used a 95 % confidence interval for all statistical tests. We found that the soil water samples taken at forested or non-forested sites were never significantly different and thus merged these data.

We investigated the sources of streamflow using two- and three-component mixing analyses and investigated the difference between the observed solute concentrations and those estimated assuming linear mixing of baseflow and rainfall. Ideally, we would use the soil water and groundwater samples taken directly before the rainfall events, but these data are not available. Instead, we have data from sampling campaigns 2 to 9 d before (event II) or after the events (I, III and

Table 2. Average concentrations (± standard deviation) for all groundwater (GW_{avg}; $n = 335$), all riparian groundwater (G1; $n = 99$), all hillslope groundwater (G2; $n = 99$), soil water (SW; $n = 116$), and rainfall samples (P; $n = 156$). Solutes are ordered by their respective groups (Sect. 4.3; Fig. 6).

Solute	Unit	GW_{avg}	G1	G2	SW	P
$\delta^{18}O$	‰	-11.0 ± 0.9^{b}	-10.8 ± 1.0^{ab}	-10.9 ± 1.1^{ab}	-10.4 ± 1.6^{a}	-12.3 ± 4.0^{c}
$\delta^{2}H$	‰	-76.0 ± 7.5^{b}	-74.3 ± 8.0^{ab}	-74.9 ± 9.1^{ab}	-70.8 ± 12.4^{a}	-84.4 ± 33.0^{c}
D_{ex}	‰	12.0 ± 0.8^{a}	12.4 ± 0.8^{a}	11.8 ± 0.9^{a}	12.0 ± 2.4^{a}	14.1 ± 3.2^{b}
Cl	$\mu g\,L^{-1}$	830.8 ± 1076.5^{a}	708.8 ± 570.1^{a}	890.5 ± 804.9^{a}	1070.3 ± 1026.6^{a}	327.1 ± 348.7^{b}
Zn	$\mu g\,L^{-1}$	593.9 ± 1745.7^{a}	720.4 ± 2218.7^{a}	698.5 ± 843.8^{a}	23.3 ± 12.5^{b}	19.3 ± 43.0^{b}
Cd	$\mu g\,L^{-1}$	0.05 ± 0.08^{a}	0.0 ± 0.1^{a}	0.1 ± 0.1^{b}	0.03 ± 0.06^{a}	0.1 ± 0.2^{b}
Ni	$\mu g\,L^{-1}$	3.2 ± 4.1^{d}	1.7 ± 1.4^{ab}	5.6 ± 6.6^{ac}	2.5 ± 1.5^{ad}	0.3 ± 0.3^{b}
Na	$\mu g\,L^{-1}$	1587.6 ± 2672.7^{b}	1107.1 ± 1000.8^{ab}	827.6 ± 341.3^{ac}	839.1 ± 565.0^{bc}	148.7 ± 153.5^{c}
Mg	$\mu g\,L^{-1}$	2235.7 ± 1730.3^{b}	1292.5 ± 684.3^{ab}	1164.1 ± 435.6^{ab}	13612.8 ± 10924^{c}	26.6 ± 18.9^{a}
Ca	$\mu g\,L^{-1}$	$56\,993.7 \pm 21966.1^{b}$	$44\,794.0 \pm 17097.6^{a}$	$55\,624.6 \pm 18099.0^{b}$	$22\,261.7 \pm 27287.8^{c}$	213.4 ± 202.7^{d}
Ba	$\mu g\,L^{-1}$	99.2 ± 171.6^{a}	64.2 ± 115.2^{a}	112.3 ± 258.6^{a}	37350 ± 27637^{b}	4.8 ± 11.8^{a}
Co	$\mu g\,L^{-1}$	0.8 ± 1.05^{c}	1.1 ± 1.0^{a}	0.3 ± 0.2^{b}	0.9 ± 1.1^{ac}	0.02 ± 0.02^{b}
Cu	$\mu g\,L^{-1}$	64.9 ± 143.7^{c}	7.4 ± 16.1^{a}	175.5 ± 211.8^{b}	5.2 ± 9.0^{a}	1.4 ± 1.0^{a}
SO_4	$\mu g\,L^{-1}$	3600.0 ± 5112.5^{b}	2511.6 ± 2843.2^{ab}	2418.7 ± 1848.2^{ab}	1602.0 ± 3061.9^{a}	623.1 ± 980.1^{c}
K	$\mu g\,L^{-1}$	530.1 ± 428.0^{c}	328.3 ± 219.2^{a}	670.3 ± 543.4^{bc}	754.1 ± 970.8^{b}	92.2 ± 91.9^{d}
Fe	$\mu g\,L^{-1}$	390.7 ± 1271.1^{a}	608.3 ± 1648.4^{a}	25.4 ± 38.6^{b}	254.3 ± 775.9^{ab}	3.5 ± 7.1^{b}
Mn	$\mu g\,L^{-1}$	592.4 ± 1111.6^{c}	1007.8 ± 911.3^{a}	68.4 ± 100.5^{b}	139.9 ± 326.2^{b}	1.3 ± 1.4^{b}

Different superscript letters ($^{a-d}$) indicate significantly different average concentrations.

IV). Since the spatial variability in groundwater composition in the Studibach is larger than the temporal variability (Kiewiet et al., 2019), we assume that the groundwater and soil water samples reflect the typical composition and variability of soil water and groundwater but acknowledge that absolute concentrations might have been slightly different. A principal component analysis (PCA) on the chemical and isotopic composition of all groundwater ($n = 335$) and soil water ($n = 116$) samples (z transformed) showed that soil water and groundwater were consistently different in the principal component space; only six of the soil water samples (5 %) plotted within the same area as the groundwater samples (see Fig. S1 in the Supplement for the PCA result and Table 2 for the average concentrations).

We estimated the fraction of event (f_e) and pre-event (f_{pe}) water in the stream water samples (C_t) using two-component isotope hydrograph separation (Eq. 2). The results for δ^2H and $\delta^{18}O$ were similar (difference between the event average $f_{pe} \leq 0.05$). Because the ratio of precision to range was better for δ^2H, we report only the δ^2H results. A pre-event baseflow sample was used to characterise the pre-event water composition (C_{pe}). The incremental weighted mean of rainfall was used to characterise the event water composition (C_e).

$$f_{pe} = \frac{C_t - C_e}{C_{pe} - C_e}. \quad (2)$$

We also estimated the fractions of groundwater, soil water and rainwater in each stream water sample using a three-component end-member mixing analysis (EMMA; Christophersen and Hooper, 1992). We based the EMMA on the first two principal components of a PCA that included all conservative tracers. We considered a tracer conservative if the concentration was linearly correlated to that of at least one other tracer (cf. Barthold et al., 2011). To determine the conservativeness, we used all groundwater, soil water and stream water samples used in this study ($n = 549$) and set the threshold for a linear correlation to $R^2 \geq 0.5$ and $p < 0.01$. Electrical conductivity (EC), calcium, magnesium, barium, δ^2H and $\delta^{18}O$ were conservative based on this definition; the other tracers (e.g. copper, sulfate, potassium and iron) were not. However, note that this threshold does not per se imply a linear trend, and that although a linear trend is consistent with conservative mixing, it does not necessarily confirm conservative mixing either (James and Roulet, 2006).

We used a Gaussian error propagation method (Genereux, 1998) to estimate the uncertainty in the calculated fractions of the source waters for the two-component hydrograph separation and EMMA. For the two-component hydrograph separation, we defined the uncertainty in the event and pre-event water composition as the standard deviation of the rainfall sampled during the event and groundwater sampled during the snapshot campaign closest to the event (see Table 1), respectively. For the uncertainty in the EMMA, we used the standard deviation of groundwater, soil water and rainwater samples for the event. We used the laboratory accuracy for the uncertainty of the stream water samples in the two-component hydrograph separation, and for the EMMA we

assumed that the uncertainty for the stream water samples in the principal component space was similar to the standard deviation of the last three stream water samples taken during each event (i.e. the last streamflow samples taken at the falling limb of the hydrograph). We multiplied the standard deviation with a t value based on the number of samples and used a 95 % confidence interval for all uncertainty estimations.

3.4.3 Deviation of concentrations from mixing of baseflow and rainfall

We compared the measured streamflow concentrations for each solute to the concentration that would be expected based on conservative mixing of rainfall and baseflow (C_{es}) as follows:

$$C_{\mathrm{es\_}x} = \left(C_{\mathrm{BF\_}x} \cdot f_{\mathrm{pe}}\right) + \left(C_{\mathrm{P\_}x} \cdot \left(1 - f_{\mathrm{pe}}\right)\right), \quad (3)$$

where $C_{\mathrm{es\_}x}$ is the "estimated" concentration for solute x, $C_{\mathrm{BF\_}x}$ and $C_{\mathrm{P\_}x}$ are the concentrations for solute x in baseflow and rainfall (average rainfall composition – Table 2), and f_{pe} is the pre-event water fraction for that sample as determined from the two-component hydrograph separation using δ^2H as the tracer (Eq. 2).

We compared the estimated ($C_{\mathrm{es\_}x}$) and measured streamflow (C_{Q_x}) concentrations for each sample and solute to assess the relationship between discharge and the potential contribution of different source areas. We assumed that underestimation of the concentrations ($C_{Q\_x}/C_{\mathrm{es\_}x} > 1$) indicates either a contribution from source areas that have a higher concentration than the sources that contributed to baseflow or reactive transport. Similarly, overestimation of the concentrations ($C_{Q\_x}/C_{\mathrm{es\_}x} < 1$) indicates either a contribution from source areas that did not contribute during baseflow and have a lower concentration than the sources that contributed to baseflow or reactive transport. Given the characteristic concentrations in different (ground-)water types (Tables 2 and 3; Fig. 2), we interpret the changes in the stream water composition during an event as follows: (1) higher copper and nickel concentrations are indicative of contributions from hillslopes and forested areas, (2) higher iron and manganese concentrations are indicative of contributions from riparian areas, (3) higher D_{ex}, barium, and magnesium concentrations are indicative of soil water, and (4) higher potassium concentrations can indicate either soil water or hillslopes groundwater. However, note that the variability for soil water, groundwater and rainfall was large (Table 2; see Fig. S2 for boxplots of tracer concentrations in each water compartment). Also, the non-conservative nature of these tracers should be taken into account. For instance, iron and manganese are mainly soluble under anoxic, reducing conditions, such as in the riparian areas, but they might oxidise and form an insoluble compound after entering the streams. Adsorption of metals (e.g. iron, copper and zinc) to organic compounds or clay particles may also influence the concentrations in streamflow,

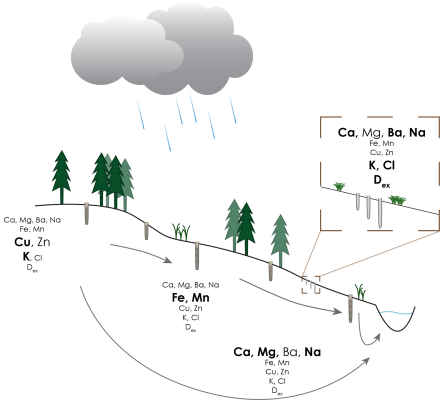

Figure 2. Illustration of a hillslope cross section with different (ground-)water compartments (based on Kiewiet et al., 2019 and Table 2), showing the tracers used in combination with δ^2H and δ^{18}O to characterise the different source areas. For most elements, the concentrations were low in rainfall compared to the concentrations in the other water compartments. High potassium, barium and chloride concentrations and high deuterium excess (D_{ex}) are indicative of soil water. For shallow groundwater, the concentrations of copper and potassium were higher at (forested) ridge locations, whereas for sites with water tables that are persistently close to the surface, the concentrations of iron and manganese were higher. We assume that higher concentrations of geogenic solutes (calcium, magnesium and sodium) indicate longer subsurface residence times. The isotopic composition for the different water compartments depends on the composition of recent precipitation.

and their concentration may be underestimated if they are adsorbed to coarser particles that settle out during streamflow recession (Kaushal et al., 2018). The concentration of some solutes is, furthermore, controlled by weathering processes or influenced by plant uptake because they are macro (potassium and magnesium) or micro (e.g. copper and nickel) plant nutrients. In this study, we assume that concentration increases or decreases due to weathering or plant uptake are negligible at the event (i.e. hourly) timescale.

3.4.4 Groundwater-level-based connectivity assessment

We investigated into how far stream chemistry reflects conservative mixing of baseflow and precipitation and whether this breaks down at a certain discharge or reflects an increase in hydrologic connectivity. We related the ratio of the measured and estimated concentrations ($C_{Q\_x}/C_{\mathrm{es\_}x}$, see Sect. 3.4.3) for each solute to the discharge and the calculated fraction of the catchment that was connected to the stream. We used the data-driven model of Rinderer et al. (2019) to

Table 3. Summary of the groups of the solutes (A–D, based on the relative concentrations during all four events; Fig. 6; NG indicates that this solute is not assigned to a group), the typical response of solute concentrations to increasing discharge (++: strong enrichment, mean $R_x > 1.5$; +: enrichment, mean R_x between 1 and 1.5; –: dilution, and mean $R_x < 1$; ±: mixed response), and ratios between the average concentrations in soil water (C_{SW}) and groundwater (C_{GWavg}) as well as the groundwater from riparian wells (C_{G1}) and hillslope wells (C_{G2}; see Table 2). See Figs. 5 and 6 for example concentration–discharge relationships for each group of solutes. The solutes are sorted according to their typical response.

Solute	Group	Typical (C) response to increasing Q	C_{SW}/C_{GWavg}	C_{G2}/C_{G1}
D_{ex}	NG	++	1	1
Cl	NG	++	1.3	1.3
Fe	D	+	0.7	∼ 0
Mn	D	+	0.2	0.1
Co	C	±	1.1	0.3
Cu	C	±	0.1	23.7
SO_4	C	±	0.4	1
K	C	±	1.4	2
Cd	A	±	0.6	–
Zn	A	±	∼ 0	1
Ni	NG	±	0.8	3.3
Na	B	–	0.5	0.7
Mg	B	–	6.1	0.9
Ca	B	–	0.4	1.2
Ba	B	–	376.5	1.7

determine which parts of the catchment were active and connected to the stream. This model uses the water level data from all 51 wells in the catchment and time series clustering to assign each pixel in the catchment to one of six groundwater level clusters based on topography. For each time step, the average relative groundwater level for all monitoring wells that belong to a cluster is calculated and assigned to all pixels in that cluster. This relative water level is then transformed into an absolute water level based on the correlation between soil depth and slope. If this simulated water level is within 30 cm of the soil surface (i.e. the part of the soil where the hydraulic conductivity is high), the pixel is considered active; otherwise, it is considered inactive. If a pixel is active and, based on surface topography, connected to the stream via other active pixels, it is assumed to be connected to the stream. We thus assume that significant lateral flow occurs when the water table rises into the near-surface layers where the hydraulic conductivity is much larger (cf. Schneider et al., 2014). Hence, the simulated connectivity refers to the connectivity of groundwater flow in the more permeable layer of the soil above the more permanently saturated soil. In the Studibach, there is an almost permanent water table in the low-conductivity gleysols in most locations. It is thus not so likely that the lateral water flow would infiltrate into the bedrock before reaching the stream (Jackson et al., 2014). Rinderer et al. (2019) tested the sensitivity of the method for misclassification of the clusters by randomly reassigning pixels to different clusters and the uncertainty in the soil depth by comparing the connectivity time series to the time series computed with a different (DEM-based) soil depth map. The soil depth had only a minor influence on the model results (RMSE > 0.0003 % of the relative soil depth). Still, misclassification of pixels (i.e. assigning them to a different cluster) could result in a difference of up to 8 % in the simulated connected area between the different model runs.

4 Results

4.1 Event characteristics

Total rainfall for the four events ranged between 17 and 33 mm (Table 1; Fig. 3). The duration of the events ranged from 7 to 28 h. The four events were larger than the long-term average daily precipitation and within the upper 30th percentile of daily precipitation at the long-term meteorological station Erlenhöhe, located 500 m from the catchment outlet (median – 10.0 mm; mean ± SD – 14.1 ± 13.8 mm for all 7452 d with more than 1 mm of precipitation between 1981–2017; Stähli, 2018). However, the events were smaller than the 50 mm threshold for large contributions of event water to streamflow (Fischer et al., 2017). The average and maximum 10 min rainfall intensities ranged between 1.2 and 3.9 mm h^{-1} and between 7 and 24 mm h^{-1}, respectively.

Discharge at the catchment outlet increased the least (from 0.02 to 0.07 mm h^{-1}) for the smallest event (I) and most for event III (0.08 to 0.43 mm h^{-1}). The simulated fraction of the catchment that was hydrologically connected to the stream varied from 0.27 (before the start of events I and II) to 0.68 (at the time of peak flow for event III; Fig. 4). The relationship between the simulated fraction of the catchment that was connected to the stream and discharge was non-linear for all events (Fig. 5, top row). For all of the four events, connectivity was lower on the rising limb of the hydrograph than on the falling limb for the same discharge. For event I, the connected area increased significantly at the recession of the streamflow. For event II, connectivity increased little during the sampling period (0.27 to 0.28). Discharge increased to > 4 mm h^{-1} after the sampling period of event II due to additional rainfall, but interestingly the simulated connectivity increased only marginally (up to 0.35; see Fig. S3) during this period. During the smaller events with initially low connectivity, the hydrologically connected area extended laterally from the stream up but remained confined to the flat areas. For the intermediate events (III and IV), the lateral extension was larger and parts of the hillslopes became connected. However, the data-based model suggested that during all four events, large parts of the catchment remained hydrologically disconnected from the stream network (Table 1; Fig. 4).

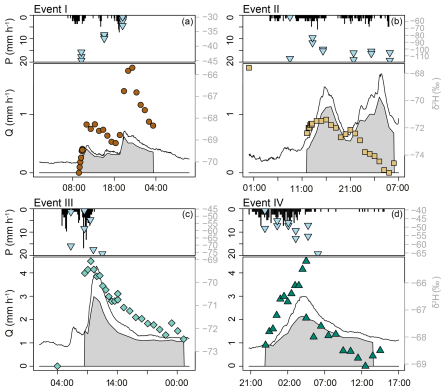

Figure 3. Hydrographs and hyetographs for the four studied events (I–IV). For each event, the upper panel shows the 10 min rainfall intensity (mm h^{-1}; bar graph) and the isotopic composition of the rainfall (δ^2H in ‰; light blue triangles), while the lower panel shows the discharge at the catchment outlet (mm h^{-1}; solid line), the isotopic composition of stream water (δ^2H in ‰; brown dots, light brown squares, turquoise diamonds and green triangles for events I–IV, respectively), and the pre-event water fraction of streamflow based on two-component hydrograph separation using δ^2H (grey polygon) as a tracer.

4.2 Concentration–discharge relationships

The chemical and isotopic composition of stream water changed during all four events, but the magnitude and direction of the response differed for each event and solute (Fig. 5). Hysteresis in the relationship between solute concentrations and discharge depended on the event size and differed between solutes (Table 3; Fig. 5). During events III and IV, the relationship between discharge and concentration was hysteretic for most solutes. The double discharge peaks during events I and II (Fig. 2) resulted in a double loop in the concentration–discharge relationship for deuterium, iron and calcium (Fig. 5).

The average relative concentration (average R_x for the streamflow samples taken during the four events, $n = 100$; Eq. 1) for deuterium excess (D_{ex}) and chloride was 4.1 and 2.0, respectively. This reflects the substantial increase in these concentrations during events. Manganese and iron concentrations also increased with increasing discharge but less than D_{ex} and chloride (mean R_x – 1.0 for both iron and manganese; maximum R_x – 2.8 and 3.2, respectively). On average, the concentrations of copper, nickel and zinc decreased with increasing discharge (mean R_x – 0.78, 0.63 and 0.31), but individual stormflow samples were enriched up to 1.7, 1.3 and 1.1 times the baseflow concentration, respectively.

Table 4. Event-average pre-event water fraction (f_{pe}) based on the two-component hydrograph separation using δ^2H as a tracer, and the event-average fractions of groundwater (f_{GW}), soil water (f_{SW}) and rain water (f_P) based on the three-component end-member mixing analyses and the associated uncertainties for both calculations.

Event	Two component		Three-component end-member mixing analyses			
	f_{pe}	Uncertainty	f_{GW}	f_{SW}	f_P	Uncertainty
I	0.86	0.28	0.81	~ 0	0.19	0.16
II	0.76	0.61	0.49	0.27	0.24	0.14
III	0.81	0.69	0.72	0.01	0.27	0.16
IV	0.78	0.25	0.74	0.01	0.25	0.14

Concentrations of iron and copper were higher on the falling limb than on the rising limb (anticlockwise hysteresis). Event I was the only event during which copper concentrations did not increase with increasing discharge.

The concentrations of sodium, magnesium, calcium and barium decreased with increasing discharge (mean R_x – < 0.77). The concentrations of these solutes, and also sulfate, were higher on the rising limb than on the falling limb (resulting in clockwise hysteresis). Sulfate concentrations decreased with increasing discharge during events I, III and IV but increased with discharge during event II. Potassium and sulfate concentrations (range R_x – 0.2–1.7 and 0.3–1.4, respectively) were highest shortly after the onset of an event (first four samples) and decreased afterwards. These differences in the magnitude and timing of the change in solute concentrations and isotopic composition allowed for subdivision of the tracers into four different groups based on the computed R_x values for all events (A to D; Table 3; Fig. 6).

4.3 Hydrograph separation and end-member mixing analysis results

Two-component hydrograph separation indicated that most stormflow was "old" water (Fig. 3; Table 3). The maximum event water fraction (f_e) was highest for event II ($f_e = 0.24 \pm 0.61$) and lowest for event IV ($f_e = 0.14 \pm 0.28$). However, the differences between the events were much smaller than the associated uncertainties (Table 4). The high event water fraction of event II occurred when the connected area was relatively small. The fraction of connected area during event II expanded by only 0.01 (up to 0.28) during the period that we sampled (see Fig. S3).

It was possible to calculate the relative fractions of groundwater, soil water and rainwater in stormflow for all events based on EMMA as well (Table 4). Groundwater dominated streamflow during all events (range f_{GW} – 0.49 ± 0.14 to 0.81 ± 0.19). The event-average soil water fraction was considerable during event II (f_{SW} – 0.27) but negligible during the other events (f_{SW} – ~ 0). The event-average pre-event water fractions based on the EMMA (i.e. the sum of the

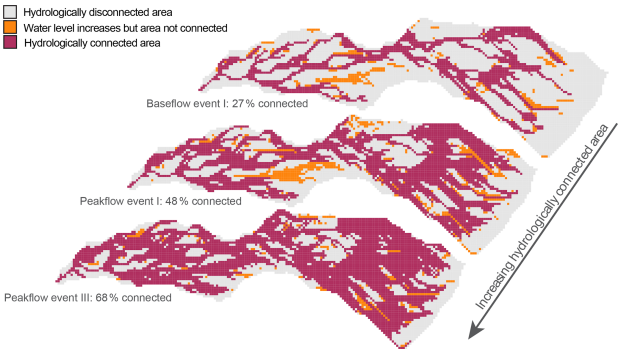

Figure 4. The simulated hydrologically connected area for three different flow conditions: from relatively low flow (baseflow prior to event I; top) and intermediate flow conditions (peak flow during event I; middle) to the period of highest discharge for the studied events (peak flow during event III; bottom). Grey indicates the hydrologically disconnected areas (water level more than 30 cm from the soil surface), red indicates the hydrologically connected area (i.e. water level within 30 cm from the soil surface and connected to the stream via other active areas), and orange indicates the active but disconnected area (i.e. the water level increased into the upper 30 cm of the soil but is not connected to the stream network by other active areas). The connected area was simulated based on the measured groundwater levels and a data-driven model that uses surface topography to estimate the water level for unmonitored grid cells (cf. Rinderer et al., 2019).

groundwater and soil water fractions) were similar to the pre-event water fractions estimated using δ^2H as a tracer in the two-component hydrograph separations (range $f_{GW} + f_{SW}$ – 0.73 to 0.81 versus range f_{pe} – 0.76 to 0.86). Although the results were similar, the uncertainties for EMMA were smaller than for the two-component hydrograph separation. The uncertainties for the EMMA results were mainly caused by the uncertainty in the groundwater fraction (contribution of the groundwater uncertainty to the total uncertainty – 97 %, 50 %, 94 % and 94 % for events I–IV, respectively). This is due to the large contribution of groundwater to streamflow and the large spatial variability in the groundwater composition. For event II, the uncertainty due to the soil water contributions was larger than for the other events (25 % for event II versus 0.01 %, 3 % and 5 % for event I, III and IV, respectively).

The explanatory power of the first two principal components for all stormflow, soil water and groundwater samples was 76.3 % for event I (PC1 – 53.1 %; PC2 – 23.2 %) and 82.0 % for event III (PC1 – 56.2 %; PC2 – 25.8 %; Fig. 7a and c). For event II and IV the explanatory power was 72.6 % and 83.8 %, respectively (see Fig. S4). The most striking aspect of the mixing plots, however, is the small change in the composition of stormflow compared to the spatial variation in the composition of the soil and groundwater end members (Fig. 7b and d). The observed changes in solute concentrations in streamflow were largest during event II (e.g. changes of 23 µg L^{-1} for Ba; 39 mg L^{-1} for Ca and 11 ‰ for δ^2H) but this change was similar to or smaller than the standard deviation of the concentrations for the groundwater samples or soil water samples taken during the corresponding snapshot campaign (e.g. groundwater – 44 µg L^{-1} for Ba, 27 mg L^{-1} for Ca and 5.9 ‰ for δ^2H; soil water – 22 310 µg L^{-1} for Ba, 23 mg L^{-1} for Ca and 10.4 ‰ for δ^2H; see Fig. S2 for boxplots of the concentrations for the different water types).

4.4 Estimated solute concentrations based on conservative mixing of rainfall and baseflow

The concentrations estimated based on the assumption of conservative mixing between rainfall and baseflow (C_{es}; Eq. 3) differed from the measured stormflow concentrations (C_Q) for almost all solutes (Fig. 8). The measured concentrations for geogenic solutes (shown for calcium and sodium in Fig. 8) were lower than the estimated concentrations. The measured concentrations of sulfate were lower than estimated based on conservative mixing as well, except for event II. For potassium there was no clear pattern; the concentrations were underestimated and overestimated at both low and high discharge (Fig. 8). The measured concentrations of cobalt, copper, nickel and iron (solute groups A and C; see Fig. 6) were slightly lower than the estimated concentrations for low discharge but (much) higher during high discharge (Fig. 8). There was no distinct threshold in the relationship between C_Q/C_{es} and either the discharge or the simulated fraction of the catchment that was connected to the stream (Figs. 8 and S5); C_Q/C_{es} rather changed gradually with increasing discharge and connected area.

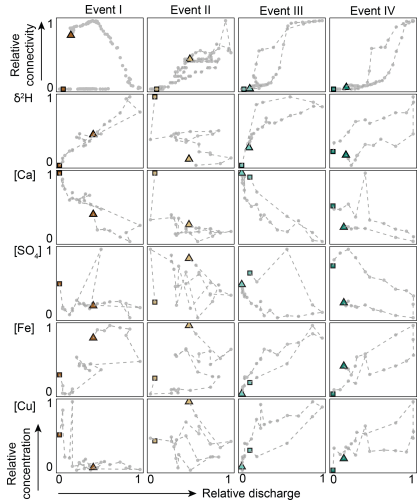

Figure 5. Relationship between the fraction of the catchment that was connected (relative connectivity) and discharge (top row), and concentration–discharge relationships for δ^2H, calcium, sulfate, iron and copper (rows 2–6) for events I–IV (columns). Individual samples are marked with a grey dot and connected with a broken line. The first sample of the event is indicated by a square and the last sample by a triangle. All data are normalised between 0 (minimum measured value for the event) and 1 (maximum measured value for the event) for better visualisation of the hysteretic relationship.

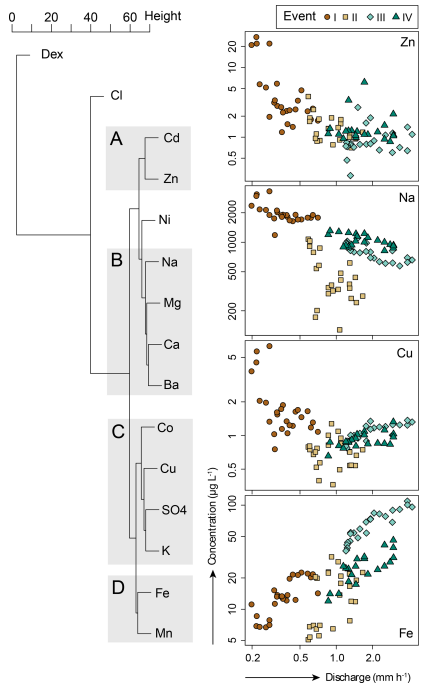

Figure 6. Dendrogram for the hierarchical clustering of solutes and D_{ex} based on the magnitude and timing of changes in streamflow concentrations compared to the baseflow concentration (R_x; Eq. 2) during the four events (I–IV), and concentration–discharge relationships for one solute from each group (A–D).

5 Discussion

5.1 Small changes in streamflow composition compared to the spatial variability in groundwater and soil water

Changes in solute concentrations in stream water during rainfall events depend on the changes in the relative contributions of different sources to streamflow (e.g. event and pre-event water or different pre-event water sources), the differences in the concentrations of these sources, as well as reactive transport processes. Our results show that the change in streamflow composition during the four rainfall events was much smaller than the spatial variability in groundwater and soil water composition. For instance, the average change in the concentration of barium and deuterium in streamflow for the four events was similar to the spatial variability in shallow groundwater and soil water measured after events I and II (13.8 µg L^{-1} Ba and 6.1 ‰ change in stream water, versus an interquartile range of 30 µg L^{-1} and 4.8 ‰ for shallow groundwater and 10.6 mg L^{-1} and 5.7 ‰ in soil water). This was also evident from the principal component analysis and mixing plots (Fig. 7). It is to be expected that the change in stream water composition is less than the variability between the end members, but for a viable hydrograph separation, the change in stream water composition should be larger than the variability within the end members (Hooper, 2001). The change in stream water composition during the four events presented in this study was not large enough to distinguish contributions from the different groundwater sources, although it is evident that pre-event water dominated streamflow.

We could show that the spatial variation within different source areas was large compared to the temporal variation because we collected a large dataset of groundwater and soil water samples. However, in other small catchment studies,

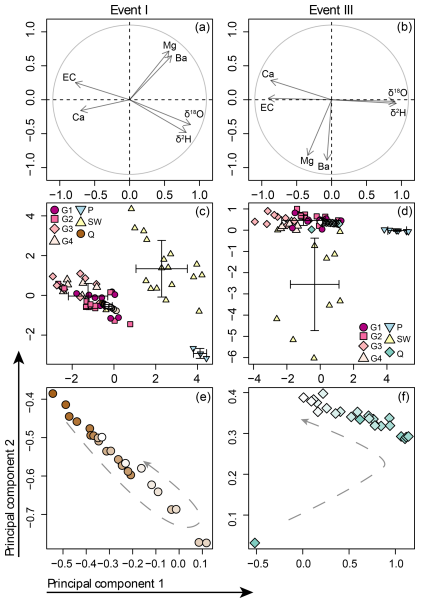

Figure 7. PCA results and mixing diagrams for events I (small event) and III (intermediate-sized event). In the biplots **(a)** and **(b)**, the length of the arrow represents the explanatory power. The mixing diagrams based on the first two principal components (**c** and **d**) show the individual rainfall (light blue triangles), soil water (yellow triangles) and groundwater samples (purple circles, pink squares, light pink diamonds and rose triangles representing groundwater types 1–4; based on Kiewiet et al., 2019), and the streamflow (Q) samples and average and standard deviation for each component (error bars). Panels **(e)** and **(f)** show an enlargement of the streamflow samples and highlight the evolution of the stream water composition (colours fade to white towards the end of the event); the general direction of change is indicated with a grey arrow and broken lines. The biplots and mixing plots for the events II and IV are shown in Fig. S4.

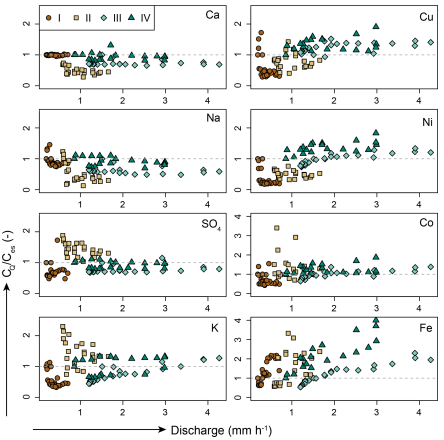

Figure 8. The ratio of the measured (C_Q) and estimated stormflow concentrations (C_{es}; Eq. 3) for calcium, sodium, sulfate, potassium, cobalt, copper, nickel and iron as a function of discharge at the catchment outlet. The broken grey line indicates where C_Q and C_{es} are equal; the different symbols reflect the different events (I–IV). Note the difference in scale for cobalt and iron. For the relationship with the simulated fraction of the catchment that was connected to the stream, see Fig. S5.

this comparison is often restricted because of insufficient spatial sampling (Penna and van Meerveld, 2019). Based on our experience for the Studibach, we see a clear need for further spatial sampling of groundwater and soil water in other catchments to determine this spatial variability.

5.2 Which areas or sources contribute to stormflow?

For the events included in this study, the estimated area that was hydrologically connected to the stream never fell to levels below a quarter of the catchment area, increased laterally upslope from the stream and increased to a maximum of two-thirds of the catchment area. The simulated connected area during a relatively small event (event I; total rainfall 17 mm) increased by a fifth of the catchment area, which implies that even small rainfall events can activate a sizable part of the catchment. The connectivity simulations for event II, however, suggest that during long-duration, low-intensity rainfall events, the change in connectivity can be small. For this event, the relative contributions of soil water and rainfall to stormflow were much higher than for the other events (Table 4).

Using a combination of different tracers to identify the sources of streamflow can be helpful because it enhances the likelihood that sources that contribute little to stormflow are identified (Barthold et al., 2017) and thereby reduces the risk of false conclusions about catchment functioning (Barthold et al., 2011). For instance, McCallum et al. (2012) used differential flow gauging and conservative (Cl) and non-conservative (Rn and EC) tracers to quantify the inflows and outflows of groundwater along three ~ 30 km long stream reaches in the Cockburn River, Logan River and Nambucca River catchments ($> 400\,\mathrm{km}^2$) in southeastern Australia. They found that predictions made with flow data alone varied significantly from predictions that also included

tracer data, and that the use of multiple tracers reduced the error in the calculation of the groundwater contributions. Moreover, the discrepancy between the results of source area analyses based on conservative and non-conservative tracers are hypothesised to indicate when sources other than baseflow and rainfall contribute to stream water (Kirchner, 2003). We found that the event water fractions from two-component hydrograph separation (isotopes) and EMMA (multi-tracer) were comparable (Table 4). Similar to our results, Ladouche et al. (2001) found for the 0.8 km^2 Strengbach catchment in France that the hydrograph separation results based on $\delta^{18}O$ (f_{pe} – 10 %) were relatively similar to the results of their mixing analyses (including DOC, Si, Ba and U), and that a multi-tracer approach allowed them to distinguish between pre-event water contributions from the upper and lower part of the catchment. We found that concentrations of metals, such as iron or copper, were much higher than expected from mixing of rainfall and baseflow, whereas weathering-derived solutes, such as sodium or calcium, were lower than expected from mixing of rainfall and baseflow. We assume that the differences between measured and expected concentrations, particularly on the falling limb and at peak flow, are at least partly caused by contributions from groundwater sources or soil water (particularly for event II) that did not contribute to baseflow (see Table 3 for ratios of concentrations in different source waters). For instance, the differences for weathering-derived solutes could be due to contributions from soil water which has lower concentrations of these solutes than groundwater. The concentrations of iron increased throughout the event until peak flow and were higher on the falling limb than on the rising limb. Since riparian groundwater has relatively high concentrations of iron (Tables 2 and 3), contributions from riparian-like areas that did not contribute to baseflow (such as flatter areas away from the stream network) during rainfall events could explain this increase. Measured copper concentrations were much higher than expected for events III and IV but lower than expected for most samples of events I and II. Because copper concentrations are relatively high for hillslope groundwater and low in soil water (Table 2 and 3; Kiewiet et al., 2019), this could be an indication that the hillslopes did not actively contribute to streamflow during events I and II and were only activated after peak flow for events III and IV (see wide hysteresis for event I in Fig. 5, top row). However, the copper concentrations should then also not have increased compared to baseflow during event II, which was not the case (maximum R_{Cu} during event II – 1.7 versus 1.0, 1.0 and 1.4 during event I, III and IV, respectively). The potassium concentrations were too variable to aid further interpretation, which is probably due to the high variation in potassium concentrations in soil water and groundwater (Table 2).

The contribution from soil water was considerable (f_{SW} – 0.27) for only one of the four events (event II; Table 4). This was a long, low-intensity event occurring on a relatively "dry" catchment (baseflow event I and II – 0.2 mm h^{-1} versus 0.7 mm h^{-1} for event III and IV). Hagedorn et al. (2000), analysed three rainfall events (7, 8 and 30 mm) in the neighbouring Erlenbach catchment and showed a large contribution of soil water to streamflow. Their mixing diagrams using chloride and calcium indicate that the average contribution of the top soil to streamflow was larger than 50 %. However, chloride and calcium concentrations vary considerably in both soil and groundwater (average coefficient of variation – 0.86 and 1.0 for eight soil water ($n = 6$ to 18) and 1.0 and 0.3 for nine groundwater ($n = 34$ to 47) snapshot campaigns for chloride and calcium, respectively). Furthermore, the concentration of bivalent cations, like calcium, in rainwater can increase during transport through the canopy (Lindberg et al., 1986). Van Meerveld et al. (2018) showed that calcium concentrations in overland flow from small landslide areas in the Studibach were much higher than for other solutes, indicating rapid dissolution as well. The much lower soil water contributions found for this study compared to Hagedorn et al. (2000) may thus be partly caused by the choice of the tracers. Understanding the role of soil water for runoff generation is challenging because of the spatial variation in its amount (e.g. McMillan and Srinivasan, 2015), the horizontal and vertical spatial variation in soil water chemistry (Gottselig et al., 2016), and the importance of preferential flow (e.g. Wiekenkamp et al., 2016). Antecedent soil moisture conditions also affect runoff amounts and stream chemistry (Zehe et al., 2010; Uber et al., 2018; Knapp et al., 2020) as well as hillslope–stream connectivity (Penna et al., 2011). Further investigation of the response of soil water, the distribution of soil water chemistry and the interaction between soil water and groundwater during rainfall events is thus important if we want to understand the influence of soil water on hydrologic connectivity and when and where soil water contributes to streamflow.

The typically moderate event water fractions could indicate that overland flow is of minor importance for streamflow in the Studibach. However, overland flow does occur in the Studibach (van Meerveld et al., 2018). Saturation overland flow has been observed during sprinkling events for other sites on gleysols in Switzerland as well (Feyen et al., 1996; Weiler et al., 1999; Badoux et al., 2006). Given the low event water fractions, we suspect that the overland flow mixes with pre-event soil water on its way to the stream (Kienzler and Naef, 2008; Elsenbeer and Vertessy, 2000) or originates from exfiltrating soil water or groundwater and thus does not have the same composition as rainwater (Barthold et al., 2017). Alternatively, overland flow may infiltrate in unsaturated soils before reaching the stream and thus not influence the stream water composition.

5.3 Hydrologic connectivity and stream water chemistry

The simulations of the active and connected area suggest that the near-stream areas are most often connected and respond

first to rainfall, highlighting their importance for the rapid generation of streamflow. The model results also showed that some areas remain disconnected from the stream (Fig. 4). Nippgen et al. (2015) found very similar connectivity patterns for a subcatchment of the Tenderfoot Creek Experimental Forest (5.55 km^2) in central Montana, USA. They simulated the connected area over a 2 year period and found that it expanded from areas parallel to and close to the stream during low-flow conditions and to the hillslopes during high-flow conditions, and that 10 % of the catchment was never connected to the stream.

The change in stream water chemistry also suggests that the connected area increased rapidly because, even for small increases in discharge, stormflow could not be described as a mixture of rainfall and baseflow. However, there was no clear relationship between the extent of the hydrologically connected area and the discrepancy between the relative changes in the concentrations of conservative and non-conservative solutes (Fig. S5). Other studies that used stream water chemistry to investigate hydrological connectivity focused on one tracer that was clearly different for different source areas (e.g. Soulsby et al., 2007; Ocampo et al., 2006). These studies illustrated that for some catchments the changes in stream water chemistry reflect changes in hydrological connectivity. However, other studies showed that the interpretation of stream-based measurements may not always be straightforward because the changes in stream water chemistry can be obscured by dampening and mixing processes (Tetzlaff et al., 2014), or because a tracer might only reflect connectivity to a specific part of the catchment rather than catchment-wide connectivity (e.g. areas with high-DOC concentrations for Pacific et al., 2010). For instance, Pacific et al. (2010) compared changes in stream water DOC concentrations with estimates of upslope riparian–stream (URS) connectivity (methods cf. Jencso et al., 2009) in the Tenderfoot Creek catchment. They found a negative (though insignificant) relationship between stream DOC export and URS connectivity and showed that URS connectivity is particularly important for predicting DOC export when areas with high-DOC concentrations are connected to the stream. Multiple studies in the Girnock catchment in Scotland used stream water Gran alkalinity and isotopic composition to investigate hydrologic connectivity (Soulsby et al., 2007; Tetzlaff et al., 2014). Birkel et al. (2010), furthermore, explored the catchment's functioning with a spatially and temporally dynamic saturation model. These studies found that contributions from the upper soil layers and upslope areas dominated streamflow at higher flows, and that there was a soil moisture threshold for the contribution of these sources (Birkel et al., 2010). Furthermore, Tetzlaff et al. (2014) showed that the dynamic behaviour of the isotopic composition of stream water was in the range of the composition of soil water from the riparian peat soils at 10 and 30 cm deep and only deviated from this range during some larger events. They concluded from these results that precipitation inputs drive the dynamics of streamflow and stream water isotopic composition, but that the streamflow responses are dampened because the water travels through different hydropedological units.

Despite substantial changes in the hydrologically connected area and the large spatial variability in groundwater composition, we did not observe a distinct threshold in the relationship between the deviation of stream chemistry from conservative mixing of rainfall and baseflow and streamflow or the connected area. The gradual change in stream water chemistry might reflect the gradual increase in the connected area with increasing discharge for all of the studied events, except event I, for which the connectivity increased abruptly after peak discharge (top row in Fig. 5). Abbott et al. (2018) showed that changes in stream water composition with increasing discharge and connectivity are less pronounced for catchments with a myriad of source areas than for catchments with fewer different landscape elements. The Studibach is characterised by many small landscape elements, particularly steep hillslopes and flatter wet areas, which formed due to landslides and soil creep and which induce small-scale differences in drainage and thus soil and vegetation development. Hence, activation of different landscape elements might occur gradually and at many different places across the catchment (i.e. the connected area extends from flat locations to the hillslopes at many different locations), but these elements all have a slightly different chemical composition. From this perspective, it is perhaps not surprising that solute concentrations in stormflow changed little compared to the spatial variability in the end-member composition because streamflow is a mixture of the many different water sources in a catchment.

Alternatively, the simulations of the active and connected areas might overestimate the change in the source areas compared to reality. Although most flow occurs in the upper, more permeable layer of the soil, seepage to deeper soil layers (Feyen et al., 1999), or to the bedrock in areas where there is no continuous groundwater table in the Gleysol, may have limited the downslope travel distance (cf. Jackson et al., 2014). We did not consider a limitation of the downslope travel distance due to bedrock infiltration because the occurrence of a permanent water table in a large part of the catchment implies that percolation to the bedrock is very slow. However, bedrock infiltration might occur at some locations (e.g. the more densely rooted forested sections on steeper, better-drained soils) and might decrease the lateral distance that a water parcel can travel. Additionally, we did not consider an offset in the timing of the simulated connectivity and response in stream water chemistry due to the travel time to the stream or mixing of hillslope and riparian groundwater in the riparian zone. Chanat and Hornberger (2003) showed with a virtual experiment for a 10 km^2 hypothetical catchment that the change in the chemical signature of the stream water can be delayed relative to the change in discharge, and that this delay was larger when the near-stream reservoir (i.e. riparian zone) was larger. Their findings are thus especially

important to consider for "wet" catchments that have a large near-stream reservoir or for which the near-stream reservoir expands quickly. Furthermore, the stormflow composition is the result of mixing of contributions from different source areas. Subsurface mixing can result in temporally variable end-member compositions. Frameworks to handle time-variable end-member compositions exist (Harris et al., 1995), but there are obvious challenges, such as measuring these time-variable compositions. Furthermore, mixing of different water sources will dampen the tracer signal (Abbott et al., 2018; Tetzlaff et al., 2014) or may even chemically "reset" the hillslope signal as it mixes with riparian groundwater (Tetzlaff et al., 2014; Lidman et al., 2017).

6 Conclusions

The results of this study show that the spatial variability in soil water and groundwater compositions across the small, pre-Alpine headwater study catchment was large. Hydrograph separation and EMMA indicated that pre-event groundwater was the dominant source of streamflow, and that soil water contributions were minimal for three of the four events. For most solutes, the stream water concentrations could not be explained by conservative mixing of baseflow and rainfall. The differences were largest at high discharges. This suggests that this deviation may indicate the contribution from new contributing sources due to the expansion of the connected area. Concentrations of weathering-derived solutes decreased more than expected, which might be due to the contributions of soil water. In contrast, concentrations of iron and copper increased more than expected, which might be due to contributions from riparian-like areas and hillslopes, respectively. Thus, the differences between the expected and measured concentrations could be partly explained by contributions from other source areas. However, there was no threshold in the relationship between streamflow and the deviations of the measured concentrations and expected concentrations based on conservative mixing, suggesting that there was no sudden activation and connection of source areas. The lack of a threshold relationship between the deviations in the solute concentrations and streamflow made it more difficult to infer changes in hydrological connectivity from the stream water solute concentrations. Overall, this work shows that inferring hydrological connectivity from solute concentrations is not straightforward, especially if we consider the large variability of the tracer concentrations in the different water sources. The gradual changes in stream water chemistry during events are likely the result of increases in the contributions from many (small) landscape elements in the catchment and reflect the gradual increase in hydrologic connectivity.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/hess-24-3381-2020-supplement.

Author contributions. LK and IvM conceptualised the study. LK collected and analysed the data and prepared the first draft of the manuscript. IvM, JS and MS provided recommendations for the data analysis, participated in discussions about the results, and edited and commented on the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work would not have been possible without the help and support of many people in the field and lab. We particularly thank Michael Rinderer and Benjamin Fischer for the initial installation of the wells, weirs and flumes; Michael Rinderer for running the data-based connectivity model for the Studibach catchment; Barbara Herbstritt for the isotope analyses; and Bjorn Studer for the cation and anion analyses. We thank the editor and two anonymous reviewers for their helpful comments to improve the paper, and the Oberallmeindkorporation Schwyz (OAK), the municipality of Alpthal and the Department of Environment of the Canton of Schwyz for the excellent co-operation.

Review statement. This paper was edited by Pilar Llorens and reviewed by two anonymous referees.

References

Abbott, B. W., Gruau, G., Zarnetske, J. P., Moatar, F., Barbe, L., Thomas, Z., Fovet, O., Kolbe, T., Gu, S., Pierson-Wickmann, A. C., Davy, P., and Pinay, G.: Unexpected spatial stability of water chemistry in headwater stream networks, Ecol. Lett., 21, 296–308, https://doi.org/10.1111/ele.12897, 2018.

Allaire, S. E., Sylvain, C., Lange, S. F., Thériault, G., and Lafrance, P.: Potential Efficiency of Riparian Vegetated Buffer Strips in Intercepting Soluble Compounds in the Presence of Subsurface Preferential Flows, PloS one, 10, 1–21, https://doi.org/10.1371/journal.pone.0131840, 2015.

Badoux, A., Witzig J., Germann P. F., Kienholz H., Lüscher P., Weingartner R., and Hegg C.: Investigations on the runoff generation at the profile and plot scales, Swiss Emmental, Hydrol. Process., 20, 377–394, https://doi.org/10.1002/hyp.6056, 2006.

Barthold, F. K., Tyralla, C., Schneider, K., Vaché, K. B., Frede, H. G., and Breuer, L.: How many tracers do we need for end member mixing analysis (EMMA)? A sensitivity analysis, Water Resour. Res., 47, 1–14, https://doi.org/10.1029/2011WR010604, 2011.

Barthold, F. K., Turner, B. L., Elsenbeer, H., and Zimmermann, A.: A hydrochemical approach to quantify the role of return flow in a surface flow-dominated catchment, Hydrol. Process., 31, 1018–1033, https://doi.org/10.1002/hyp.11083, 2017.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrolog. Sci. J., 24, 43–69, https://doi.org/10.1080/02626667909491834, 1979.

Birkel, C., Tetzlaff, D., Dunn, S. M., and Soulsby, C.: Towards a simple dynamic process conceptualization in rainfall–runoff models using multi-criteria calibration and tracers in temperate, upland catchments, Hydrol. Process., 24, 260–275, https://doi.org/10.1002/hyp.7478, 2010.

Blume, T. and van Meerveld, H. J. I.: From hillslope to stream: methods to investigate subsurface connectivity, WIREs Water, 2, 177–198 https://doi.org/10.1002/wat2.1071, 2015.

Bracken, L. J. and Croke, J.: The concept of hydrological connectivity and its contribution to understanding runoff-dominated geomorphic systems, Hydrol. Process., 21, 2267–2274, https://doi.org/10.1002/hyp.6313, 2007.

Brown, V. A., McDonnell, J. J., Burns, D. A., and Kendall, C.: The role of event water, a rapid shallow flow component, and catchment size in summer stormflow, J. Hydrol., 217, 171–190, https://doi.org/10.1016/S0022-1694(98)00247-9, 1999.

Burns, D. A., Hooper, R. P., McDonnell, J. J., Freer, J. E., Kendall, C., and Beven, K.: Base cation concentrations in subsurface flow from a forested hillslope: The role of flushing frequency, Water Resour. Res. 34, 3535–3544, https://doi.org/10.1029/98WR02450, 1998.

Buttle, J. M.: Isotope hydrograph separations and rapid delivery of pre-event water from drainage basins, Prog. Phys. Geogr. 18, 16–41, https://doi.org/10.1177/030913339401800102, 1994.

Chanat, J. G. and Hornberger, G. M.: Modeling catchment-scale mixing in the near-stream zone – Implications for chemical and isotopic hydrograph separation, Geophys. Res. Lett., 30, 1091, https://doi.org/10.1029/2002GL016265, 2003.

Christophersen, N. and Hooper, R. P.: Multivariate Analysis of Stream Water Chemical Data' The Use of Principal Components Analysis for the End-Member Mixing Problem, Water Resour. Res., 28, 99–107, https://doi.org/10.1029/91WR02518, 1992.

Detty, J. M. and McGuire, K. J.: Threshold changes in storm runoff generation at a till – mantled headwater catchment, Water. Resour. Res., 46, 1–15, https://doi.org/10.1029/2009WR008102, 2010.

Devito, K. J. and Hill, A. R.: Sulphate dynamics in relation to groundwater – Surface water interactions in headwater wetlands of the southern Canadian Shield, Hydrol. Process., 11, 485–500, https://doi.org/10.1002/(SICI)1099-1085(199704)11:5<485::AID-HYP455>3.0.CO;2-F, 1997.

Elsenbeer, H. and Vertessy, R. A.: Stormflow generation and flowpath characteristics in an Amazonian rainforest catchment, Hydrol. Process., 14, 2367–2381, https://doi.org/10.1002/1099-1085(20001015)14:14<2367::AID-HYP107>3.0.CO;2-H, 2000.

Evans, C. and Davies, T. D.: Causes of concentration/discharge hysteresis and its potential as a tool for analysis of episode hydrochemistry, Water Resour. Res., 34, 129–137, https://doi.org/10.1029/97WR01881, 1998.

Feyen, H., Leuenberger, J., Papritz, A., Gysi, M., Flühler, H., and Schleppi P.: Runoff processes in catchments with a small scale topography, Phys. Chem. Earth., 21, 177–181, https://doi.org/10.1016/S0079-1946(97)85581-4, 1996.

Feyen, H., Wunderli, H., Wydler, H., and Papritz, A.: A tracer experiment to study flow paths of water in a forest soil, J. Hydrol., 225, 155–167, https://doi.org/10.1016/S0022-1694(99)00159-6, 1999.

Fischer, B. M. C., Rinderer, M., Schneider, P., Ewen, T., and Seibert, J.: Contributing sources to baseflow in pre-alpine headwaters using spatial snapshot sampling, Hydrol. Process., 29, 5321–5336, https://doi.org/10.1002/hyp.10529, 2015.

Fischer, B. M. C., Stähli, M., and Seibert, J.: Pre-event water contributions to runoff events of different magnitude in pre-alpine headwaters, Hydrol. Res., 48, 28–47, https://doi.org/10.2166/nh.2016.176, 2017.

Fischer, B. M. C., Aemisegger, F., Graf, P., Sodemann, H., and Seibert, J.: Assessing the Sampling Quality of a Low-Tech Low-Budget Volume-Based Rainfall Sampler for Stable Isotope Analysis, Front. Earth Sci., 7, 1–8, https://doi.org/10.3389/feart.2019.00244, 2019.

Genereux, D.: Quantifying uncertainty in tracer-based hydrograph separations, Water Resour. Res., 34, 915–919, https://doi.org/10.1029/98WR00010, 1998.

Godsey, S. E., Kirchner, J. W., and Clow, D. W.: Concentration – discharge relationships reflect chemostatic characteristics of US catchments, Hydrol. Process., 23, 1844–1864, https://doi.org/10.1002/hyp.7315, 2009.

Gottselig, N., Wiekenkamp, I., Weihermüller, L., Brüggemann, N., Berns, A. E., Bogena, H. R., Borchard, N., Klumpp, E., Lücke, A., Missong, A., Pütz, T., Vereecken, H., Huisman, J. A., and Bol, R.: A Three-Dimensional View on Soil Biogeochemistry: A Dataset for a Forested Headwater Catchment, J. Environ. Qual., 46, 210–218, https://doi.org/10.2134/jeq2016.07.0276, 2016.

Hagedorn, F., Schleppi, P., Waldner, P., and Flühler, H.: Export of dissolved organic carbon and nitrogen from Gleysol dominated catchments–the significance of water flow paths, Biogeochem., 50, 137–161, https://doi.org/10.1023/A:1006398105953, 2000.

Harris, D. M., McDonnell, J. J., and Rodhe, A.: Hydrograph Separation Using Continuous Open System Isotope Mixing, Water Resour. Res., 31, 157–171, https://doi.org/10.1029/94WR01966, 1995.

Hooper, R. P.: Applying the scientific method to small catchment studies: a review of the Panola Mountain experience, Hydrol Process., 15, 2039–2050, https://doi.org/10.1002/hyp.255, 2001.

Hooper, R. P. and Shoemaker, C. A.: A Comparison of Chemical and Isotopic Hydrograph Separation, Water Resour. Res., 22, 1444–1454, https://doi.org/10.1029/WR022i010p01444, 1986.

Hooper, R. P., Christophersen, N., and Peters, N. E.: Modelling streamwater chemistry as a mixture of soilwater end-members – An application to the Panola Mountain catchment, Georgia, U.S.A., J. Hydrol., 116, 321–343, https://doi.org/10.1016/0022-1694(90)90131-G, 1990.

Hopp, L. and McDonnell, J. J.: Connectivity at the hillslope scale: Identifying interactions between storm size, bedrock permeability, slope angle and soil depth, J. Hydrol., 376, 378–391, https://doi.org/10.1016/j.jhydrol.2009.07.047, 2009.

Hornberger, G. M., Scanlon, T. M., and Raffensperger, J. P.: Modelling transport of dissolved silica in a forested headwater catchment: the effect of hydrological and chemical time scales on hys-

teresis in the concentration – discharge relationship, Hydrol. Process., 15, 2029–2038, https://doi.org/10.1002/hyp.254, 2001.

Jackson, C. R., Bitew, M., and Du, E.: When interflow also percolates: downslope travel distances and hillslope process zones, Hydrol. Process., 28, 3195–3200, https://doi.org/10.1002/hyp.10158, 2014.

James, A. L. and Roulet, N. T.: Investigating the applicability of end-member mixing analysis (EMMA) across scale: A study of eight small, nested catchments in a temperate forested watershed, Water Resour. Res., 42, W08434, https://doi.org/10.1029/2005WR004419, 2006.

Jencso, K. G. and McGlynn, B. L.: Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation, Water Resour. Res., 47, 1–16, https://doi.org/10.1029/2011WR010666, 2011.

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., and Marshall, L. A.: Hydrologic connectivity between landscapes and streams: transferring reach and plot scale understanding to the catchment scale, Water Resour. Res., 45, W04428, https://doi.org/10.1029/2008WR007225, 2009.

Kaushal, S. S. Gold, A. J., Bernal, S., Newcomer Johnson, T. A., Addy, K., Burgin, A., Burns, D. A., Coble, A. A., Hood, E., Lu, Y., Mayer, P., Minor, E. C., Schroth, A. W., Vidon, P., Wilson, H., Xenopoulos, M. A., Doody, T., Galella, J. G., Goodling, P., Haviland, K., Haq, S., Wessel, B., Wood, K. L., Jaworski, N., and Belt, K. T.: Watershed "chemical cocktails": forming novel elemental combinations in Anthropocene fresh waters, Biogeochem., 141, 281–305, https://doi.org/10.1007/s10533-018-0502-6, 2018.

Kennedy, V. C., Zellweger, G. W., and Avanzino, R. J.: Variation of rain chemistry during storms at two sites in northern California, Water Resour. Res., 15, 687–702, https://doi.org/10.1029/WR015i003p00687, 1979.

Kienzler, P. M. and Naef, F.: Temporal variability of subsurface stormflow formation, Hydrol. Earth Syst. Sci., 12, 257–265, https://doi.org/10.5194/hess-12-257-2008, 2008.

Kiewiet, L., von Freyberg, J., and van Meerveld, H. J. I.: Spatiotemporal variability in hydrochemistry of shallow groundwater in a small pre-alpine catchment: The importance of landscape elements, Hydrol. Process., 33, 2502–2522, https://doi.org/10.1002/hyp.13517, 2019.

Kiewiet, L., van Meerveld, I., and Seibert, J.: Effects of spatial variability in the groundwater isotopic composition on hydrograph separation results for a pre-alpine catchment, Water Resour. Res., 56, e2019WR026855, https://doi.org/10.1029/2019WR026855, 2020.

Kirchner, J. W.: A double paradox in catchment hydrology and geochemistry, Hydrol Process., 17, 871–874 https://doi.org/10.1002/hyp.5108, 2003.

Knapp, J. L. A., von Freyberg, J., Studer, B., Kiewiet, L., and Kirchner, J. W.: Concentration–discharge relationships vary among hydrological events, reflecting differences in event characteristics, Hydrol. Earth Syst. Sci., 24, 2561–2576, https://doi.org/10.5194/hess-24-2561-2020, 2020.

Ladouche, B., Probst, A., Viville, D., Idir, S., Baqué, D., Loubet, M., Probst, J. L., and Bariac, T.: Hydrograph separation using isotopic, chemical and hydrological approaches (Strengbach catchment, France), J. Hydrol., 242, 255–274, https://doi.org/10.1016/S0022-1694(00)00391-7, 2001.

Landwehr, J. M. and Coplen, T. B.: Line-conditioned excess: A new method for characterizing stable hydrogen and oxygen isotope ratios in hydrologic systems (IAEA-CSP–26/P), International Atomic Energy Agency (IAEA), available at: https://inis.iaea.org/search/search.aspx?orig_q=RN:37043527 (last access: June 2020), 2006.

Lehmann, P., Hinz, C., McGrath, G., Tromp-van Meerveld, H. J., and McDonnell, J. J.: Rainfall threshold for hillslope outflow: an emergent property of flow pathway connectivity, Hydrol. Earth Syst. Sci., 11, 1047–1063, https://doi.org/10.5194/hess-11-1047-2007, 2007.

Lidman, F., Boily, Å., Laudon, H., and Köhler, S. J.: From soil water to surface water – how the riparian zone controls element transport from a boreal forest to a stream, Biogeosciences, 14, 3001–3014, https://doi.org/10.5194/bg-14-3001-2017, 2017.

Lindberg S. E., Lovett G. M., Richter D. D., and Johnson D. W., Atmospheric Deposition and Canopy Interactions of Major Ions in a Forest, Science, 231, 141–145, https://doi.org/10.1126/science.231.4734.141, 1986.

McCallum, J. L., Cook, P. G., Brunner, P., Berhane, D., Rumpf, C., and McMahon, G. A., Quantifying groundwater flows to streams using differential flow gaugings and water chemistry, J. Hydrol., 416–417, 118–132, https://doi.org/10.1016/j.jhydrol.2011.11.040, 2012.

McGlynn, B. L. and McDonnell, J. J.: Quantifying the relative contributions of riparian and hillslope zones to catchment runoff, Water Resour. Res., 39, 1310, https://doi.org/10.1029/2003WR002091, 2003.

McMillan, H. K. and Srinivasan, M. S.: Characteristics and controls of variability in soil moisture and groundwater in a headwater catchment, Hydrol. Earth Syst. Sci., 19, 1767–1786, https://doi.org/10.5194/hess-19-1767-2015, 2015.

Nippgen, F., McGlynn, B. L., and Emanuel, R. E.: The spatial and temporal evolution of contributing areas, Water Resour. Res., 51, 4550–4573, https://doi.org/10.1002/2014WR016719, 2015.

Ocampo, C. J., Sivapalan, M., and Oldham, C.: Hydrological connectivity of upland-riparian zones in agricultural catchments: Implications for runoff generation and nitrate transport, J. Hydrol., 331, 643–658, https://doi.org/10.1016/j.jhydrol.2006.06.010, 2006.

Oswald, C. J., Richardson, M. C., and Branfireun, B. A.: Water storage dynamics and runoff response of a boreal Shield headwater catchment, Hydrol. Process., 25, 3042–3060, https://doi.org/10.1002/hyp.8036, 2011.

Pacific, V. J., Jensco, K. G., and McGlynn, B. L.: Variable flushing mechanisms and landscape structure control stream DOC export during snowmelt in a set of nested catchments, Biogeochem., 99, 193–211, https://doi.org/10.1007/s10533-009-9401-1, 2010.

Penna, D. and van Meerveld, H. J.: Spatial variability in the isotopic composition of water in small catchments and its effect on hydrograph separation, WIREs Water, 2019, e1367, https://doi.org/10.1002/wat2.1367, 2019.

Penna, D., Tromp-van Meerveld, H. J., Gobbi, A., Borga, M., and Dalla Fontana, G.: The influence of soil moisture on threshold runoff generation processes in an alpine headwater catchment, Hydrol. Earth Syst. Sci., 15, 689–702, https://doi.org/10.5194/hess-15-689-2011, 2011.

R Core Team: A language and environment for computing. R Foundation for Statistical Computing, Vienna, Austria, available at: http://www.r-project.org/ (last access: 5 May 2020), 2013.

Rinderer, M., van Meerveld, H. J., and Seibert, J.: Topographic controls on shallow groundwater levels in a steep, prealpine catchment, Water Resour. Res., 50, 6067–6080, https://doi.org/10.1002/2013WR015009, 2014.

Rinderer, M., van Meerveld, I., Stähli, M., and Seibert, J.: Is groundwater response timing in a pre-alpine catchment controlled more by topography or by rainfall?, Hydrol. Process., 30, 1036–1051, https://doi.org/10.1002/hyp.10634, 2015.

Rinderer, M., van Meerveld, H. J. I., and McGlynn, B.: From Points to Patterns: Using Groundwater Time Series Clustering to Investigate Subsurface Hydrological Connectivity and Runoff Source Area Dynamics, Water Resour. Res. 55, 1–23, https://doi.org/10.1029/2018WR023886, 2019.

Schneider, P., Pool, S., Strouhal, L., and Seibert, J.: True colors – experimental identification of hydrological processes at a hillslope prone to slide, Hydrol. Earth Syst. Sci., 18, 875–892, https://doi.org/10.5194/hess-18-875-2014, 2014.

Seibert, J., Grabs, T., Köhler, S., Laudon, H., Winterdahl, M., and Bishop, K.: Linking soil- and stream-water chemistry based on a Riparian Flow-Concentration Integration Model, Hydrol. Earth Syst. Sci., 13, 2287–2297, https://doi.org/10.5194/hess-13-2287-2009, 2009.

Soulsby, C., Tetzlaff, D., van den Bedem, N., Malcolm, I. A., Bacon, P. J., and Youngson, A. F.: Inferring groundwater influences on surface water in montane catchments from hydrochemical surveys of springs and streamwaters, J. Hydrol., 333, 199–213, https://doi.org/10.1016/j.jhydrol.2006.08.016, 2007.

Stähli, M.: Longterm hydrological observatory Alptal (central Switzerland), available at: https://www.envidat.ch/dataset/longterm-hydrological-observatory-alptal-central-switzerland (last access: 16 August 2019), 2018.

Stähli, M. and Gustafsson, D.: Long-term investigations of the snow cover in a subalpine semi-forested catchment, Hydrol. Process., 20, 411–428, https://doi.org/10.1002/hyp.6058, 2006.

Stieglitz, M., Shaman, J., McNamara, J., Engel, V., Shanley, J., and Kling, G. W.: An approach to understanding hydrologic connectivity on the hillslope and the implications for nutrient transport, Global Biogeochem. Cy., 17, 1105, https://doi.org/10.1029/2003GB002041, 2003.

Tetzlaff, D., Birkel, C., Dick, J., Geris, J., and Soulsby, C.: Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions, Water Resour. Res., 50, 969–985, https://doi.org/10.1002/2013WR014147, 2014.

Uber, M., Vandervaere, J.-P., Zin, I., Braud, I., Heistermann, M., Legoût, C., Molinié, G., and Nord, G.: How does initial soil moisture influence the hydrological response? A case study from southern France, Hydrol. Earth Syst. Sci., 22, 6127–6146, https://doi.org/10.5194/hess-22-6127-2018, 2018.

Uhlenbrook, S., Roser, S., and Tilch, N.: Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model, J. Hydrol., 291, 278–296, https://doi.org/10.1016/j.jhydrol.2003.12.038, 2004.

von Freyberg, J., Studer, B., Rinderer, M., and Kirchner, J. W.: Studying catchment storm response using event- and pre-event-water volumes as fractions of precipitation rather than discharge, Hydrol. Earth Syst. Sci., 22, 5847–5865, https://doi.org/10.5194/hess-22-5847-2018, 2018.

van Meerveld, H. J., Seibert, J., and Peters, N. E.: Hillslope-riparian-stream connectivity and flow directions at the Panola Mountain Research Watershed, Hydrol. Process., 29, 3556–3574, https://doi.org/10.1002/hyp.10508, 2015.

van Meerveld, H. J. I., Fischer, B. M. C., Rinderer, M., Stähli, M., and Seibert, J.: Runoff generation in a pre-alpine catchment: A discussion between a tracer and a shallow groundwater hydrologist, Cuad. Investig. Geogr., 44, 429–452, https://doi.org/10.18172/cig.3349, 2018.

Weiler, M., Scherrer, S., Naef, F., and Burlando, P.: Hydrograph separation of runoff components based on measuring hydraulic state variables, tracer experiments and weighting methods, IAHS Publications, 258, 249–255, 1999.

Wiekenkamp, I., Huisman, J. A., Bogena, H. R., Lin, H. S., and Vereecken, H.: Spatial and temporal occurrence of preferential flow in a forested headwater catchment, J. Hydrol., 534, 139–149, https://doi.org/10.1016/j.jhydrol.2015.12.050, 2016.

Zehe, E., Graeff, T., Morgner, M., Bauer, A., and Bronstert, A.: Plot and field scale soil moisture dynamics and subsurface wetness control on runoff generation in a headwater in the Ore Mountains, Hydrol. Earth Syst. Sci., 14, 873–889, https://doi.org/10.5194/hess-14-873-2010, 2010.

Zuecco, G., Penna, D., Borga, M., and van Meerveld, H. J.: A versatile index to characterize hysteresis between hydrological variables at the runoff event timescale, Hydrol. Process., 30, 1449–1466, https://doi.org/10.1002/hyp.10681, 2016.

Zuecco, G., Rinderer, M., Penna, D., Borga, M., and van Meerveld, H. J.: Quantification of subsurface hydrologic connectivity in four headwater catchments using graph theory, Sci. Total Environ., 646, 1265–1280, https://doi.org/10.1016/j.scitotenv.2018.07.269, 2019.

C: Paper III

Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR026855

Effects of Spatial Variability in the Groundwater Isotopic Composition on Hydrograph Separation Results for a Pre-Alpine Headwater Catchment

Leonie Kiewiet[1], Ilja van Meerveld[1], and Jan Seibert[1,2]

[1]Department of Geography, University of Zürich, Zürich, Switzerland, [2]Department of Aquatic Sciences and Assessment, University Uppsala, Uppsala, Sweden

Key Points:
- Baseflow did not reflect the catchment average groundwater composition
- Uncertainties in hydrograph separation are likely higher than usually reported due to spatial variability in pre-event water
- Using samples that represent the spatial variability in pre-event water composition yields more robust pre-event water fractions

Supporting Information:
- Supporting Information S1
- Supporting Information S2
- Supporting Information S3

Correspondence to:
L. Kiewiet,
leonie.kiewiet@geo.uzh.ch

Citation:
Kiewiet, L., van Meerveld, I., & Seibert, J. (2020). Effects of spatial variability in the groundwater isotopic composition on hydrograph separation results for a pre-Alpine headwater catchment. *Water Resources Research*, *56*, e2019WR026855. https://doi.org/10.1029/2019WR026855

Received 28 NOV 2019
Accepted 28 APR 2020
Accepted article online 4 MAY 2020

©2020. American Geophysical Union. All Rights Reserved.

Abstract Isotope hydrograph separation is a powerful tool to investigate catchment functioning. In most hydrograph separation studies, a pre-event baseflow sample is used to represent the pre-event water, and thus, baseflow is assumed to be a mixture of all the water that is stored in the catchment. However, baseflow may not be representative of all water stored in the catchment because some sources may not contribute to baseflow. This is problematic when the isotopic composition of the sources is highly variable. We quantified the effects of spatial variability in the shallow groundwater isotopic composition on pre-event water characterization and hydrograph separation results. We compared the composition of groundwater sampled at 38 wells in a 0.2 km^2 pre-alpine catchment with stream water sampled before, during, and after three rainfall events. We estimated the number of groundwater samples needed to characterize the average groundwater composition in the catchment and its spatial variability and compared the results of two-component hydrograph separations for different ways to characterize the pre-event water. We found that differences in the calculated pre-event water fractions and uncertainties were large and depended on which and how many samples were used to characterize the pre-event water composition. Analyses based on a limited number of groundwater samples likely underestimate the real uncertainty and can give a false impression of accuracy. Our results highlight the importance of representing the variability in the pre-event water composition when applying hydrograph separation analyses. We therefore recommend sampling pre-event water at multiple locations or estimating the variability based on literature values.

Plain language summary For prediction of floods, droughts, or water quality, it is important to understand how rainfall becomes streamflow. One question is how much rainfall contributes to streamflow immediately ("new" water) and how much of the streamflow is groundwater that has been in the catchment for some time ("old" water). One way to answer this question is to look at changes in the stream water composition. For this, the composition of the "old" water needs to be specified, for instance, by taking a stream water sample before the rain starts or a number of groundwater samples. Usually, researchers take only a few samples to determine this "old" water composition. However, the groundwater composition varies from location to location. We calculated at how many locations one has to take a groundwater sample to reliably estimate the "old" water composition. We also calculated the amounts of "new" and "old" water in streamflow for three rainfall events based on different samples to characterize the "old" water. We found that the calculated amounts were different and that using more samples provides more robust results. Thus, we should take multiple samples that represent the variability in groundwater across the entire catchment when estimating rainfall and groundwater contributions to streamflow.

1. Introduction

Groundwater is the main contributor to streamflow in undisturbed headwater catchments in temperate climates (Buttle, 1994; Klaus & McDonnell, 2013; Rodhe, 1987; Sklash & Farvolden, 1979), but the relative contribution of groundwater varies between and during events and is affected by antecedent moisture conditions and rainfall amount and intensity (e.g., Fischer et al., 2016; Penna et al., 2015; Tetzlaff et al., 2014). The two-component isotope hydrograph separation method is often used to determine the relative contributions of pre-event water (or groundwater and soil water) and event water (precipitation) to the stream. The method assumes conservative mixing of event and pre-event water. One of the main assumptions of isotope hydrograph separation is that these sources have a constant and distinctly different isotopic signature or that

Water Resources Research 10.1029/2019WR026855

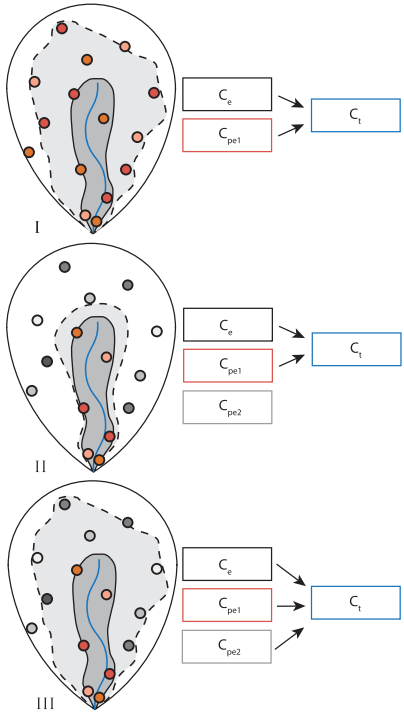

Figure 1. Schematic representation of the changes in the contributing area (white area: does not contribute to streamflow, light gray area: only contributes during intermediate or high flow conditions, dark gray area: always contributes) and the spatial variability in the groundwater composition (represented by different colored dots) for three situations, and contributions of event (C_e) and pre-event (C_{pe1} and C_{pe2}) water to streamflow (C_t) during peakflow conditions for each situation. Situation I (top): There is no significant spatial variability in the isotopic composition of the groundwater. Even though the contributing area changes during the event, the pre-event baseflow sample characterizes the pre-event water composition well. Situation II (middle): Even though there is significant spatial variability in the isotopic composition of the groundwater, the hillslopes do not contribute to streamflow and the relative contributions of the water sources do not change during the event. The pre-event baseflow sample, thus, represents the pre-event water that contributes to streamflow during the event. Situation III (bottom): There is significant spatial variability in the isotopic composition of the groundwater and the contributing area changes during the event. The pre-event baseflow sample does not adequately represent the pre-event water that contributes to streamflow during the event.

any variation in the signature can be accounted for (Buttle, 1994). For a viable hydrograph separation, the variability in the signature of the water sources should be smaller than in streamflow (Hooper, 2001). However, the spatial variability in the isotopic signature of the water sources is often not fully characterized or accounted for due to a lack of data.

Usually, we assume that pre-event streamflow is a mixture of all the water that is stored in the catchment and thus represents the catchment average pre-event water composition (C_{pe}). However, the water that is stored in a catchment can be highly variable in its isotopic composition (Kendall et al., 2001; Kiewiet et al., 2019; McDonnell et al., 2007). This spatial variability in the pre-event water composition does not affect hydrograph separation results as long as the relative contributions of the different water sources to streamflow during the event do not differ from the pre-event contributions (Figure 1, Situation II). However, not all parts of the catchment are hydrologically connected to the stream during baseflow conditions (Jencso et al., 2010; Jencso & McGlynn, 2011), and the relative contributions from different groundwater (and soil water) stores change with the expansion of the contributing area and connection of different source areas (Rinderer et al., 2019) (Figure 1, Situation III). For example, McGlynn and McDonnell (2003) found that between events, throughout small events, and in the early part of large events, streamflow consisted mainly of riparian groundwater in a 2.5-ha, steep headwater catchment in New Zealand. However, for large events, the contribution of hillslope runoff was similar to that of riparian groundwater. Similarly, Jencso et al. (2010) showed that the composition of stream water in the moderately sloping Tenderfoot Creek Experimental Forest in the United States (range subcatchment sizes: 3–22.8 km^2) shifted toward the hillslope signature once a groundwater connection was established. Other three-component hydrograph separation studies have also shown significant contributions of hillslope water during rainfall events (Burns et al., 2003; Inamdar & Mitchell, 2007; Penna et al., 2016). If hillslope groundwater has a composition that is different from riparian groundwater and contributes to the stream during an event but not to baseflow, then the pre-event baseflow sample does not accurately reflect the composition of the pre-event groundwater that contributes to streamflow during the event (Figure 1, Situation III). The effects of this difference between the sampled and the actual pre-event water composition on hydrograph separation results have been highlighted by modeling studies (e.g., McCallum et al., 2010; Jones et al., 2006) but have not been quantified with field data on the variability in the isotopic composition of groundwater.

A theoretic calculation, assuming that streamflow is a mixture of riparian groundwater, hillslope groundwater, and precipitation but that pre-event streamflow (i.e., baseflow) is fed only by riparian groundwater, provides insight in the associated error in the event water fraction (see supporting information S1). For example, if the difference in the isotopic composition of riparian and hillslope groundwater is 2‰ (with the hilllslope pre-event water source being more depleted than the riparian water that contributes to baseflow) and the precipitation is 5‰ enriched compared to baseflow, then the error in the calculated event water fraction is 8% when hillslope pre-event water contributes 20% to streamflow, but the error is 20% when the hillslope water contributes 50% of the streamflow. However, it is unlikely that there are only two sources of pre-event water in a catchment and

Water Resources Research 10.1029/2019WR026855

that both sources are well mixed, as the isotopic composition of groundwater, soil water, seeps, and springs can be highly variable across the catchment. Above all, this theoretical calculation of the uncertainty assumes that we know the contribution of the pre-event water sources, which we usually do not know.

Generally, the uncertainty in tracer-based hydrograph separation studies is estimated using the Gaussian error-propagation method as, for instance, presented by Genereux (1998). In this method the uncertainty depends on the difference in the composition of the event and pre-event water and the variability in the composition of the two water sources (i.e., the spatial and temporal variability in the event and pre-event water composition). Methods to handle the temporal variability in the event water composition are well established and frequently applied (Laudon et al., 2002; McDonnell et al., 1990), but there is often a lack of data on the spatial variability (e.g., Cayuela et al., 2019; Fischer et al., 2017). Information about the spatial variability of the pre-event water composition is rarely available, and thus, the uncertainty due to this variability is not well characterized (Penna & van Meerveld, 2019). Any spatial variability will result in temporal variability if the relative contributions of the water stored in different parts of the catchment change during an event. This is problematic because the total uncertainty in hydrograph separation results is most sensitive to the uncertainty in the component that contributes most to streamflow (i.e., pre-event water in most undisturbed headwater catchments in temperate climates; Genereux, 1998). Due to the lack of information, the variability in the pre-event water composition in the Gaussian error-propagation method is sometimes replaced by the analytical accuracy (e.g., Cayuela et al., 2019; Jefferson et al., 2015). This significantly underestimates the total uncertainty of the hydrograph separation results because the spatial variability is likely much larger than the analytical uncertainty and can even be the largest source of uncertainty (Uhlenbrook & Hoeg, 2003).

An alternative way to assess the uncertainty of hydrograph separation results is to perform multiple hydrograph separation calculations in which the (pre-)event water composition varies over the observed or estimated range. McDonnell et al. (1991) used this approach and showed that a ±1‰ variability in δ^2H of the pre-event water resulted in a ±10% variability in the calculated pre-event water fraction (f_{pe}) for the Maimai M8 catchment in New Zealand (i.e., $f_{pe} \pm 0.10$). Similarly, Rodhe (1981) showed that a 0.5‰ variability in $\delta^{18}O$ resulted in a ±15% variability in the pre-event water fraction (i.e., $f_{pe} \pm 0.15$) for the boreal Stormyra and Nåsten basins in northern Sweden. Measurements in other catchments indicate that the range in the isotopic composition of groundwater can be as large as the ranges assigned by Rodhe (1981) and McDonnell et al. (1991) but can also be much larger (and admittedly also smaller). For instance, Carey and Quinton (2005) reported a range of 0.7‰ to 0.8‰ $\delta^{18}O$ for a scarcely vegetated catchment in Canada based on three campaigns in which they sampled seven groundwater wells. Klaus et al. (2015) reported a range of 1.8‰ $\delta^{18}O$ for 14 wells in three small catchments in South Carolina, USA, that were sampled monthly for eight consecutive months. Kendall et al. (2001) concluded from prestorm and poststorm sampling of soil water and groundwater at the artificial hydrohill catchment in China that the variability in $\delta^{18}O$ across the catchment and in the soil profile was about 4‰. If the variability in the pre-event water composition is larger than the assumed 0.5‰ $\delta^{18}O$ by Rodhe (1981) or the 1‰ δ^2H by McDonnell et al. (1991), then the uncertainty in the two component hydrograph separation results is likely also larger than the reported 10% to 15%.

The aim of this study was to assess how the characterization of the pre-event water composition and the spatial variability in the isotopic composition of the groundwater affect two-component hydrograph separation results. We address the sensitivity of the hydrograph separation analysis to the samples that were chosen to characterize the pre-event water composition and how the number (and choice) of groundwater samples affects the calculated pre-event water fraction and its uncertainty. For this analysis, we used groundwater data from 38 wells in the 0.2 km^2 Studibach catchment in Switzerland and stream water and rainfall data for three rainfall events. More specifically, we addressed the following research questions:

- How many wells do we need to sample to adequately represent the average isotopic composition of shallow groundwater and its variability in a small headwater catchment?
- How different are hydrograph separation results when using samples from different (combinations of) wells or the pre-event streamflow sample to characterize the pre-event water composition?
- How does the number of groundwater samples used to characterize the pre-event water composition influence the uncertainty of the hydrograph separation results?

Water Resources Research 10.1029/2019WR026855

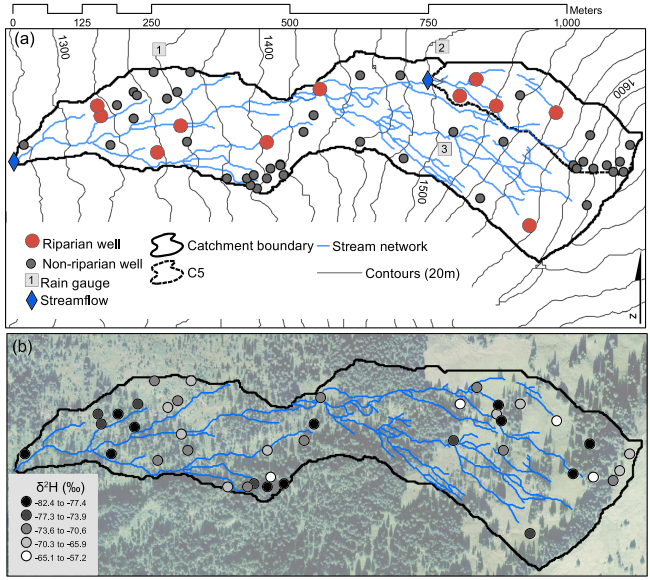

Figure 2. The Studibach with (a) the stream network (blue), 20-m elevation contour lines (gray), the catchment boundary (solid black line) and boundary of subcatchment C5 (dashed black lines) and the streamflow gauging stations (blue diamonds), groundwater wells (riparian groundwater in red circles, all other (i.e., non-riparian) wells in gray circles), and the location of the rain gauges (1–3, light gray squares). (b) δ^2H of the shallow groundwater for the 5 October 2016 snapshot campaign (color ranging from black (more depleted) to white (more enriched)) projected on an aerial photo of the catchment (source: Federal Office of Topography (Swisstopo); aerial images no. 20000090712703). See supporting information S2 for maps of the shallow groundwater isotopic composition for the other snapshot campaigns.

2. Study Site Description

The data for this study were collected in the Studibach, a 0.2 km^2 headwater catchment of the Zwäckentobel catchment, located in the Alptal, Switzerland (N°47.038, E°8.723). The catchment elevation ranges from 1,270 to 1,650 m above sea level (Figure 2). Mean annual precipitation is about 2,300 mm year^{-1} and is relatively evenly distributed throughout the year. About one third of the precipitation falls as snow (Feyen et al., 1999). During the snow-free season (June–October), it rains on average every other day (van Meerveld et al., 2018).

Soil creep and landslides have created a complex topography of very steep slopes and flatter areas. The average slope is 35° (Rinderer et al., 2014). The steeper parts of the catchment (about half of the catchment area) are covered by open coniferous forest (*Picea abies L.* with an understory of *Vaccinium* sp.) (Hagedorn et al., 2000). The flatter areas (about a third of the area) are characterized by moorlands or wet grasslands; the remaining area is covered by alpine meadows. The upper part of the catchment is used for cattle grazing in the summer. Springs and streams emerge at the transition from convex to concave slopes (Molnar et al., 2010). The stream response to rainfall is flashy, and previous studies suggest that event water fractions are generally low but highly variable. The average event water fraction for 24 events in the neighboring Erlenbach catchment ranged between 0.04 and 0.75 (median: 0.21; von Freyberg et al., 2018). The maximum event water fraction for five subcatchments throughout the Zwäckentobel catchment ranged from 0.09 to

Water Resources Research 10.1029/2019WR026855

0.90 for 13 events (Fischer et al., 2016). Event water fractions are largest for large events with wet antecedent conditions (Fischer et al., 2016; von Freyberg et al., 2018).

The flysch bedrock consists of shale, calcareous slate, and sandstone banks and is assumed to be poorly permeable (Mohn et al., 2000). The bedrock is overlain by gleysols. The soils are wet throughout most of the year. For a large part of the catchment the groundwater is close to the soil surface (Rinderer et al., 2014).

The chemistry of the shallow groundwater is dominated by the carbonate-rich bedrock and usually has a calcium-bicarbonate signature, although some sites have high magnesium and sulfate concentrations. Based on the chemical and isotopic composition, the shallow groundwater can be divided into four types, of which three represent hydrogeomorphic units (Kiewiet et al., 2019):

I. Riparian zone, near-stream areas, and other flat areas with riparian-like characteristics (referred to as "riparian" in the remainder of the text and figures) that are characterized by above average concentrations of manganese, iron, and cobalt.
II. Hillslopes and steeper areas that are characterized by above average concentrations of copper, zinc, chromium, and nickel.
III. Areas with discharge of "deep" groundwater that are characterized by higher concentrations of strontium and molybdenum.

The fourth water type was characterized by high magnesium and sulfate concentrations and likely represents flow through a specific part of the flysch. This water type was found in six wells (15% of all wells) of which five were located within 100 m from each other in a1000 m^2 subcatchment (Kiewiet et al., 2019).

The isotopic composition of the shallow groundwater in the Studibach is affected by seasonal changes in the precipitation composition and is most depleted directly after snowmelt and most enriched in late summer. Although the difference in the δ^2H values was statistically insignificant, riparian and hillslope groundwater (i.e., Types I and II) resembled the recent precipitation (i.e., these sites had a more enriched isotopic composition during the spring to fall sampling period), whereas groundwater Type III had a more depleted isotopic composition (Kiewiet et al., 2019).

There was no significant relation between the shallow groundwater isotopic composition and topography (e.g., slope, Topographic Wetness Index, distance to stream, well depth, and elevation) or hydrodynamic situation (e.g., flushing frequency and persistency of the groundwater table) (Kiewiet et al., 2019; Figure 2). The spatial variability in the groundwater isotopic composition is smallest directly after snowmelt and in late fall (Kiewiet et al., 2019). During this period, it is also most similar to baseflow (Kiewiet et al., 2019). This reflects the seasonal changes in the connectivity based on groundwater level observations between 2010 and 2014. Approximately 27% of the catchment area was continuously connected to the stream in March to June 2013, whereas only 9% of the catchment area was continuously connected in July 2013 (Rinderer et al., 2019).

3. Methods

3.1. Field Measurements

The main analyses in this study are based on stream water samples for three events (A–C) and groundwater sampled on two different dates (Table 1). We sampled one additional rainfall event for which hydrograph separation was not possible, but we use this event (D) to exemplify how the composition of baseflow and groundwater may differ.

3.1.1. Hydrometric Measurements

Rainfall was monitored at three locations with tipping-bucket rain gauges (0.2-mm resolution, Davis, Odyssey Dataflow Systems Pty Limited, New Zealand) (Figure 2a). Stream stage was monitored every 5 min at the (sub)catchment outlets (Figure 2a) with a pressure transducer (DCX-22 CTD, Keller AG für Druckmesstechnick, Switzerland). The pressure data were corrected for changes in barometric pressure using the temperature and elevation adjusted barometric pressure data from the MeteoSchweiz meteorological station in Einsiedeln (910 m above sea level; ~10 km from the catchment). The stream stage data were converted to streamflow using a stage-discharge relationship based on 16 (C5, V-notch weir) to 20 (catchment outlet, stream stage measured directly in the river) salt dilution measurements. Due to technical

Water Resources Research 10.1029/2019WR026855

Table 1
Characteristics of the Events: Total Rainfall (mm), Average and Maximum 10-min Intensity (mm h^{-1}), Duration of the Event (hr), Number of Streamflow Peaks During the Sampling Period, and Minimum (Q_{min}) and Maximum Specific Discharge (Q_{max}) During the Event (both in mm h^{-1})

			Event A	Event B	Event C
	Date event		3 October 2016	3 October 2017	5 October 2017
	Date snapshot campaign		5 October 2016 (I)	12 October 2017 (II)	12 October 2017 (II)
Rainfall	Total rainfall	mm	15a	27[a, b]	29[a, b]
	Average intensity	mm h^{-1}	1.0	3.8	1.9
	Maximum Intensity	mm h^{-1}	7.2	20.4	8.4
	Duration	hr	15	8.5	13
	Number of samples	—	10	8	11
	Total rainfall between event and snapshot campaign	mm	0.2	8.8	42
Streamflow	Number of peaks during sampling period	—	3	1	1
	Q_{min}	mm h^{-1}	0.2[c]	0.8	0.8
	Q_{max}	mm h^{-1}	0.7[c]	4.3	3.0
	Q_{max}/Q_{min}	—	3.5	5.4	3.8
	Number of streamflow samples	—	24	22	23

[a]Average values based on measurements at RG1 and RG2 [b]Average values based on measurements at RG1 and RG3 [c]Specific discharge for the catchment is estimated based on the streamflow measured at sub-catchment C5 because of missing data for the catchment outlet

issues, we do not have a complete time series of stage at the catchment outlet for Event A. We used the streamflow time series at C5 (Figure 2a) to estimate the streamflow at the outlet. More specifically, we compared the streamflow for the four months directly following Event A for both sites and used this relation to estimate the flow at the outlet for the 15-day data gap. The coefficient of determination (r^2) between specific discharge at the catchment outlet and at the C5-subcatchment outlet was 0.66; the root-mean-square error was 0.75 mm h^{-1} (the 10th and 90th percentiles of specific discharge at the catchment outlet from November to March are 0.35 and 2.11 mm h^{-1}, respectively). Although the streamflow magnitude during Event A is thus uncertain, the average pre-event water fraction would be minimally affected because the same offset would be used for the entire event.

3.1.2. Stream Water, Rainfall, and Groundwater Sampling

We used automatic samplers (full-size portable sampler, 6712, ISCO Teledyne, USA) to sample streamwater at the catchment outlet. We manually turned the samplers on when the predicted time of the onset of precipitation was during the day and used a timer if the event was predicted to start during the night. We used a multi-interval program and adjusted the sampling interval to the predicted event duration. The first six stream water samples were taken every 10 (only the shortest event: A) to 20 min; the remaining 18 samples were taken at an hourly interval.

We used sequential rainfall samplers (built after Kennedy et al. (1979), but see Fischer et al. (2019) for a detailed description) to sample the rainfall at rain gauge locations RG1 and RG3 (Figure 2a) and additionally sampled at location RG3 during Event A. The samplers function mechanically, so we calculated the time of sampling from the rainfall time series. Each sampler had 100-ml glass bottles and a collection area of 214 cm^2, which resulted in one sample for approximately every 5 mm of rainfall. The maximum number of bottles that could be filled was 12. We emptied the rainfall and stream water samplers within 24 hr after the event to avoid isotope fractionation. The samples were collected in polyethylene bottles (50 ml) and stored in the fridge (6°C) until they were transferred, within a few days, to 20-ml glass vials with a membrane screw cap.

We used the groundwater samples collected during two baseflow snapshot sampling campaigns (I and II) for the main analysis and a third sampling campaign (III) for the comparison with Event D. The snapshot campaigns are described in detail by Kiewiet et al. (2019). In short, the wells in the catchment were installed in 2009–2010 (Rinderer et al., 2014). They were hand-augered until the bedrock, screened over their entire length, except for the top 10 cm, and were sealed with a bentonite layer. The locations of the wells were based on the distribution of the Topographic Wetness Index (Beven & Kirkby, 1979), so that the wells cover the range of wet and dry locations in each subcatchment (Rinderer et al., 2014). More specifically, 8 wells are located at ridge sites, 22 at midslope sites, and 21 at footslope locations; 20 of the wells are located in forested areas and 31 in non-forested areas. Well depths range from 0.5 m at ridge sites to 2.5 m at footslope locations.

Water Resources Research 10.1029/2019WR026855

All wells were purged the day before the sampling campaign by pumping them dry or extracting at least twice the well volume. During the snapshot sampling campaigns, groundwater samples were taken from all wells that contained water (n = 34 to 38). After sample collection, the groundwater samples were treated and stored similarly as the stream water and rainwater samples. The electrical conductivity (EC) of each sample was measured in the field with a handheld device (Multi 3420, WTW GmbH, Germany), except for the stream water samples of Event A, which were measured only later in the laboratory due to equipment malfunctioning.

All stream water, rainwater, and groundwater samples were analyzed for stable water isotopes with a Cavity Ring-Down Spectroscope (L2140-I (CRDS) or L2130-I (CRDS), Picarro, Inc., USA) at the Chairs of Hydrology, University of Freiburg, Germany. The reported precision is ±0.16‰ for $\delta^{18}O$ and ±0.6‰ for $\delta^{2}H$. We calculated the line-conditioned excess (LC-excess; Landwehr and Coplen, 2006) which describes the deviation of a sample from the local meteoric water line. None of the samples deviated significantly from the local meteoric water line: the median LC excess was −2.0‰ and −2.6‰ $\delta^{2}H$ for stream water and groundwater samples, respectively. Using $\delta^{18}O$ or $\delta^{2}H$ as a tracer yielded similar hydrograph separation results. Because the ratio of precision to range (range $\delta^{18}O$: 13.98‰ and range $\delta^{2}H$: 121.9‰) was more favorable for deuterium, we only show the results for $\delta^{2}H$ here.

3.2. Data Analyses

3.2.1. Number of Samples Required to Characterize the Isotopic Composition of the Groundwater

We estimated how the number of sampled wells affects the estimates of the average and standard deviation of the isotopic composition of the groundwater in the catchment. We randomly selected a number of wells (without replacement) and calculated the average and standard deviation of $\delta^{2}H$ and EC for the selected samples. The selections were based on a constrained random approach (i.e., the values were randomly selected samples from our wells, but the locations of the wells were chosen based on their topographic characteristics; see section 3.1.2). We did this for all possible numbers of wells (1 to 38) and for all possible number of riparian wells (1 to 11). We found that the standard deviation of the average pre-event water composition for 1,000 realizations (i.e., 1,000 random selections of wells for each set of number of wells) differed less than 0.01‰ and therefore chose to limit our analysis to 1,000 realizations rather than computing all possible combinations (e.g., $3.5 \cdot 10^{10}$ combinations are possible when sampling 19 out of 38 wells; unordered random sampling $n!/(k!(n\text{-}k)!)$, where n is the total number of observations and k the number of selected observations).

We also used the basic confidence interval equation based on the normal distribution (equation 1) and rearranged it to solve for the sample size (n, equation 2) to estimate the number of wells that need to be sampled to obtain an estimate of the average isotopic composition with a 95% confidence interval. This calculation must be viewed as a rough estimation because it assumes that the samples are independent, which can be debated for shallow groundwater, and because the groundwater isotope and EC measurements were only approximately normally distributed (the Shapiro-Wilk test suggested that the data were normally distributed, p values were 0.35 and 0.41 for $\delta^{2}H$ and 0.61 and 0.55 for EC for Campaigns I and II, respectively).

$$E = t_{\alpha/2} \frac{\sigma}{\sqrt{n}} \tag{1}$$

$$n = \left(\frac{t_{\alpha/2}\sigma}{E} \right)^{2} \tag{2}$$

where n is the required number of wells, σ is the standard deviation, E is the error margin (here we use a value of 1.0‰, 1.2‰, and 2.5‰), and $t_{\alpha/2}$ is the critical value associated with a specific confidence level. $t_{\alpha/2}$ approaches 1.960 for a 95% confidence interval for sample sizes larger than 100. We used $t_{\alpha/2}$ values from the t distribution (Pearson & Hartley, 1954) when we applied this equation to small sample sizes (n < 30) (e.g., when selecting three samples, two degrees of freedom, $t_{\alpha/2}$ = 4.303).

3.2.2. Hydrograph Separation

3.2.2.1. General Approach

For the three events (A–C), we applied a two-component hydrograph separation to determine the pre-event water fraction:

Water Resources Research 10.1029/2019WR026855

$$f_{pe} = \frac{C_t - C_e}{C_{pe} - C_e} \quad (3)$$

where f_{pe} is the fraction of pre-event water, C_t the isotopic composition of stream water, and C_e and C_{pe} the isotopic composition of event water and pre-event water, respectively. For the pre-event water composition (C_{pe}), we used the pre-event baseflow sample, the baseflow sample taken during the snapshot sampling campaign, or different groundwater samples taken during the snapshot sampling campaign. We did not use any soil water samples nor did we perform three-component hydrograph separations, although soil water may also contribute to streamflow. We used the incremental weighted mean of the rainwater samples (McDonnell et al., 1990) to characterize the event water composition (C_e). We did not consider any spatial variability in the event water isotopic composition. Spatial sampling of rainfall in the Zwäckentobel suggested that the variability in event water composition does not vary significantly with elevation (Fischer et al., 2017), and our data from the two rain gauges did not suggest a significant relation either. However, the spatial variability in the event water composition can be large due to the complex topography, surrounding mountains, and forest cover, and the two rain gauges might not have been enough to capture this variability.

The two snapshot sampling campaigns used in the analyses are a subset of the nine campaigns during the snow-free seasons of 2016 and 2017 of Kiewiet et al. (2019). Ideally, the snapshot groundwater sampling campaigns would have taken place right before the sampled rainfall events. The original goal of the snapshot sampling campaigns was to determine the spatial variability in shallow groundwater composition across the catchment. The timing of these campaigns was therefore not aligned with the event sampling (which would have been logistically very challenging due to the time required to purge and sample all groundwater wells and low predictability of the moderately-sized the rainfall events in this mountainous terrain). However, two of the snapshot sampling campaigns took place shortly (2–9 days) after the sampled rainfall events (Table 1). Daily precipitation measurements and samples from the neighboring Erlenbach catchment (precipitation gauge and sampler ~500 m from the Studibach outlet; Rücker et al., 2019) suggest that there was no precipitation between the sampled events and the snapshot sampling campaigns (Event A) or there were only small, low-intensity events with an isotopic composition similar to the rainfall before the sampled event (Events B and C; see supporting information S3). We assume that these small- to medium-sized events did not significantly change the isotopic composition of the shallow groundwater because their isotopic composition was typical for the season. Furthermore, we assume that these events did not affect the spatial variability in the isotopic composition of the shallow groundwater. The observed spatial variability in the groundwater composition was large for all of the nine sampling campaigns (median standard deviation of δ^2H for all nine snapshot campaigns: 3.8‰ (range: 2.3‰ to 9.3‰; Campaign I: 5.3‰; Campaign II: 3.9‰). Thus, although the calculated pre-event water fractions might be somewhat inaccurate because the samples taken during the snapshot sampling campaigns do not perfectly represent the pre-event groundwater prior to the sampled events, we expect that this difference is small and has a negligible effect on the sensitivity and uncertainty analyses (described below).

3.2.2.2. Sensitivity to the Characterization of the Pre-Event Water Composition

To determine the sensitivity of the hydrograph separation results to the characterization of the pre-event water composition (C_{pe}), we used different (combinations of) samples (Table 2). We used the pre-event baseflow sample (BF_{pe}) and baseflow sampled during the snapshot campaign (BF_{ss}), the average isotopic composition for all groundwater wells (GW_{avg}), the average isotopic composition for all riparian wells (RP_{avg}), the composition for individual wells (GW) or individual riparian wells (RP), and subsets of three, six, or nine randomly selected groundwater wells (GW_n) or riparian wells (RP_n). Note that calculations for all groundwater wells (GW_{avg} or GW_n) also include all riparian groundwater wells. When using the isotopic composition derived from samples taken at individual wells, we repeated the hydrograph separation 11 (RP) or 38 (GW) times, that is, once for each individual well. When selecting subsets of three, six, or nine wells, we repeated the hydrograph separation calculation 1,000 times using the average composition of the randomly selected wells. For 50 realizations the standard deviation in the calculated pre-event water fraction was less than 0.1% (i.e., difference in standard deviation $f_{pe} < 0.001$). This implies that the number of 1,000 repetitions was sufficient.

AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

10.1029/2019WR026855

Table 2
Abbreviations, Description of the Sample(s) Used for each Pre-event Water Characterization, and Reason for Including the Method to Characterize the Pre-event Water Isotopic Composition (C_{pe})

Abbreviation	Sample(s) used for the pre-event water characterization	Reason for selecting this (set of) sample(s)
BF_{pe} and BF_{ss}	A baseflow sample taken before the event (BF_{pe}) or during the snapshot sampling campaign (BF_{ss}).	Baseflow is often assumed to represent the average composition of the water stored in the catchment.
RP_{avg}	The average isotopic composition of all riparian wells sampled during the snapshot campaign.	In many catchments wells are mainly located in the riparian zone or near the stream. Furthermore, this groundwater most likely contributes to baseflow.
GW_{avg}	The average isotopic composition of all wells sampled during the snapshot campaign.	These samples are assumed to represent the average composition of shallow groundwater in the catchment.
RP or GW	The composition of a sample from an individual (riparian) well.	This represents a situation where only one well is used to characterize the composition of the groundwater.
RP_n or GW_n	The average composition of three, six or nine randomly selected wells (GW_n) or three, six, or nine randomly selected riparian wells (RP_n).	This represents a situation where sampling is limited to three, six, or nine wells throughout the catchment or the riparian zone to represent the average groundwater composition.

We report the median, the minimum, and the 10th to 90th percentile range of the event-averaged pre-event water fraction for each pre-event water characterization. To determine the significance of the difference in the median event-averaged pre-event water contribution to streamflow for the different characterization methods, we performed pairwise comparisons for all combinations of characterizations (Tukey-Kramer test; Sheskin, 2003; Tukey.HSD in the "agricolae" R-package). We used a 95% confidence level for all statistical analyses.

For about 10% of the groundwater samples (different wells for each event), hydrograph separation led to physically impossible pre-event water fractions (i.e., smaller than 0 or larger than 1). We allowed a 2.5% error margin and considered all f_{pe} values outside this range (i.e., $f_{pe} < -0.025$ and $f_{pe} > 1.025$) as impossible. We excluded these values from the analysis for calculations of the minimum and event-averaged f_{pe}. For the analyses, in which we repeated the hydrograph separations 1,000 times, we set physically impossible pre-event water fractions to our lower limit or upper limit (−0.025 or 1.025) to reduce the bias induced by the exclusion of these results.

3.2.2.3. Uncertainty Estimation

We estimated the uncertainty of the calculated pre-event water fractions (W_{fpe}) using the Gaussian error-propagation method suggested by Genereux (1998):

$$W_{fpe}=\left\{\left(\frac{C_e-C_t}{(C_e-C_{pe})^2}W_{Cpe}\right)^2+\left(\frac{C_t-C_{pe}}{(C_e-C_{pe})^2}W_{Ce}\right)^2+\left(\frac{-1}{(C_e-C_{pe})^2}W_{Ct}\right)^2\right\}^{1/2} \quad (4)$$

where C_{pe}, C_t, and C_e are the isotopic composition of the pre-event water, event water, and stream water and W_{Cpe}, W_{Ce}, and W_{Ct} are the uncertainties for the pre-event water, event water, and stream water composition, respectively. We used the standard deviation of the isotopic composition of the rain samples taken during the event for W_{Ce} and the laboratory precision for W_{Ct}. We used the standard deviation of δ^2H for the groundwater samples that were used in the hydrograph separation calculation multiplied by the appropriate t value (based on the number of samples; Pearson & Hartley, 1954) for W_{Cpe}. If only one sample was used to determine C_{pe} (i.e., for BF_{pe}, BF_{ss}, RP, or GW), we used the laboratory precision for W_{Cpe} because no information on the standard deviation was available.

4. Results

4.1. Description of the Events

4.1.1. Precipitation and Streamflow

The total rainfall for Events A, B, and C was 15, 27, and 29 mm, which occurred in 15, 8.5 and 13 hr, respectively (Table 1 and Figure 3). The average and maximum 10-min rainfall intensities were 1.0, 3.8 and 1.9 mm h^{-1} and 7.8, 20.4, and 8.4 mm h^{-1}, respectively. The catchment response to rainfall was quick;

Water Resources Research 10.1029/2019WR026855

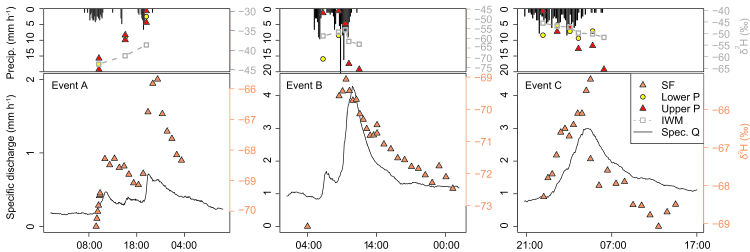

Figure 3. Time series of 10-min precipitation (bar graph, mm h^{-1}) and specific discharge at the catchment outlet (black lines, mm h^{-1}), isotopic composition (δ^2H, ‰) of streamflow (SF, orange triangles) and rainwater sampled at the lower rain gauge (RG1, lower P, yellow circles) and upper rain gauge (RG3, upper P, red triangles), and the incremental weighted mean isotopic composition of the rainwater (IWM, gray squares connected by a dashed line) for Events A (left), B (middle), and C (right). The baseflow sample for Event B was taken one day prior to the event but is projected at 04:00 for better visualization. Please note that the axes differ for the different events. See Table 1 for the dates of the events and event characteristics.

streamflow increased 3.5 (Event A) to 5.4 times (Event B) during the events (Table 1). Event A caused multiple streamflow peaks, whereas Events B and C resulted in only one peak (Table 1 and Figure 3).

4.1.2. Isotopic Composition of Rainwater

The intra-event variability in the isotopic composition of rainfall was large. The standard deviation of the δ^2H of rainwater varied from 5.2‰ for Event A ($n = 10$) to 12.6‰ for Event B ($n = 8$). For Event A, the rainwater became isotopically more enriched throughout the event, whereas during Events B and C, it became more depleted (Figure 3). During Event B, the rainwater shifted from a composition that was more enriched than stream water to a composition that was more depleted than stream water. However, the incremental weighted mean of rainwater remained 10‰ more enriched than the streamwater, so that hydrograph separation was still possible for this event.

4.1.3. Isotopic Composition of Stream Water

The isotopic composition of the stream water changed toward the composition of the precipitation during all events (Figure 3). The stream water isotopic composition changed as soon as the water level rose, but the magnitude of the response depended on the amount of rain and the difference in the isotopic composition of the rainwater and pre-event baseflow. Pre-event baseflow and the incremental weighted mean of rainwater were isotopically most similar for Event B (difference in δ^2H: 10.7‰ to 18.1‰) and most different for Event A (difference in δ^2H: 27.0‰ to 31.9‰). The stream water isotopic composition changed most during Event A (from −70.5‰ to −65.7‰) and least during Event C (from −69.1‰ to −65.2‰; Table 3 and Figure 4). The relation between the stream water isotopic composition and specific discharge varied from

Table 3
Average $\delta^2H \pm$ Standard Deviation (‰) for Event Water (C_e), Pre-event Water (C_{pe}) Based on the Sample Taken Before the Start of the Event (BF_{pe}), the Streamflow Sample Taken During the Snapshot Sampling Campaign (BF_{ss}), the Average of All Samples From Riparian Wells (RP_{avg}), and the Average of All Groundwater Samples (GW_{avg}) and Minimum and Maximum δ^2H for Streamwater ($C_{t\text{-}min}$,$C_{t\text{-}max}$) for Events A–C

Event		A	B	C
Date stormflow event	—	3 October 2016	3 October 2017	5 October 2017
Date groundwater campaign (number)	—	5 October 2016 (I)	12 October 2017 (II)	12 October 2017 (II)
Baseflow pre-event BF_{pe}	C_{pe}	−70.5	−73.8	−72.5
Baseflow snapshot campaign BF_{ss}	C_{pe}	−71.0	−72.7	−72.7
Riparian groundwater RP_{avg}	C_{pe}	−70.2 ± 4.3 (11)	−71.8 ± 3.6 (11)	−71.8 ± 3.6 (11)
All groundwater GW_{avg}	C_{pe}	−73.0 ± 5.3 (38)	−73.2 ± 3.9 (38)	−73.2 ± 3.9 (38)
Rainwater (average)	C_e	−37.0 ± 5.2 (10)	−64.6 ± 12.6 (8)	−49.2 ± 6.5 (11)
Streamwater (minimum)	$C_{t\text{-}min}$	−70.5	−73.7	−69.1
Streamwater (maximum)	$C_{t\text{-}max}$	−65.7	−69.1	−65.2

Note. The sample size is given in parentheses (n) for sample sizes larger than 1.

Water Resources Research 10.1029/2019WR026855

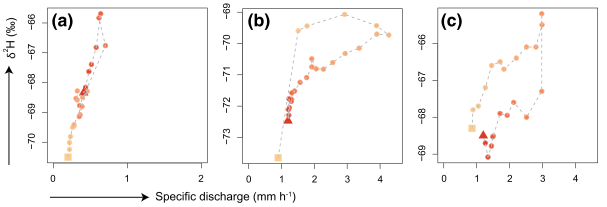

Figure 4. Relation between specific discharge (mm h^{-1}) and stream water isotopic composition (δ^2H, ‰) for Events A–C. The color of the symbols changes from light (first sample: square) to dark (last sample: triangle). The hysteresis index class (c.f. Zuecco et al., 2016) was: I for Events B and C, indicating a clockwise loop and increase from the initial concentration and IV for Event A, indicating a counterclockwise direction and increase from the initial concentration. Note that the axes differ for the events.

being rather linear (Event A) to more hysteretic (Events B and C) (Figure 4). For Event C, the isotopic composition changed markedly just after peakflow (from −65.2‰ to −67.3‰).

4.1.4. Spatial Variability in the Isotopic Composition of Groundwater

The spatial variability in the isotopic composition of the shallow groundwater was large: δ^2H varied from −86.3‰ to −67.8‰ and −80.9‰ to −57.2‰ for snapshot Campaigns I (5 October 2016) and II (12 October 2017), respectively (Table 3). Riparian groundwater (δ^2H mean ± sd: −71.8 ± 3.6‰ and −70.2 ± 4.3‰) was slightly more enriched than the catchment average groundwater (i.e., average of all sampled groundwater wells: −73.2 ± 3.9‰ and −73.0 ± 5.3‰) for both snapshot campaigns (Figure 5). This difference was larger than twice the laboratory precision (0.6‰ δ^2H) but not statistically significant. Baseflow at the catchment outlet during the snapshot campaigns (−71.0‰ and −72.7‰ δ^2H, Table 3 and Figure 5) differed 0.3‰ and 2.2‰ δ^2H from the average composition of all groundwater wells and 0.8‰ and 2.5‰ from the average composition of all riparian wells for Campaigns I and II, respectively. Pre-event baseflow differed less than 1‰ δ^2H from baseflow sampled during the snapshot campaigns.

4.2. Number of Wells Required to Characterize the Isotopic Composition of the Groundwater

The range in the calculated average isotopic composition of the groundwater decreased with an increasing number of samples (Figure 5). The 5th and 95th percentiles of the average isotopic composition of the groundwater for six randomly selected groundwater samples were −75.7‰ and −69.4‰ for Campaign I and −75.4‰ and −70.8‰ for Campaign II (Figure 5). For nine randomly selected samples, they were −75.2‰ and −70.2‰ for Campaign I and −74.9‰ and −71.3‰ for Campaign II. The difference between the 5th and 95th percentiles of the calculated average groundwater composition was less than 2.5‰ (i.e., half of the average change in the isotopic composition of stream water during the three studied events) as soon as more than 21 and 16 randomly selected samples were used to determine the average composition of the groundwater for Campaigns I and II, respectively.

The sample size calculation based on the basic confidence interval equation (equation 2) suggests that in order to obtain an estimate of the average groundwater composition within 1.2‰ δ^2H, we would have to sample 41 or 95 wells for Campaigns II and I, respectively. For an estimate of the average within 2.5‰, 12 or 24 wells need to be sampled, respectively. Adding restrictions to the random sampling scheme based on landscape characteristics (e.g., by selecting only samples from wells that are close to the stream or that have a high Topographic Wetness Index) did not yield different results.

Sometimes EC is used instead of the isotopic composition for hydrograph separation (e.g., Inserillo et al., 2017; Pellerin et al., 2008). In our campaigns, the spatial variability in EC was even larger than the variability in the isotopic composition, with a mean ± standard deviation of 443 ± 100 and 414 ± 130 μS cm^{-1} for Campaigns I and II, respectively. The average change in stream water EC was 98 μS cm^{-1} (range: 55 to 167 μS cm^{-1}). For an estimate of the average EC of groundwater within half of the change in stream water EC with 95% confidence (i.e., an error smaller than 50 μS cm^{-1}), we would need 18 or 28 samples for Campaigns I and II, respectively.

Water Resources Research 10.1029/2019WR026855

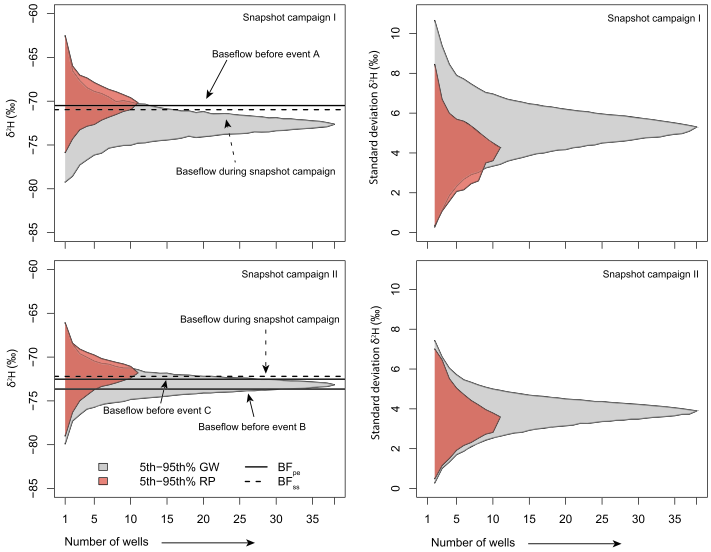

Figure 5. The 5th to 95th percentile of the average (left column) and standard deviation (right column) of the isotopic composition (δ^2H) as a function of the number of randomly selected groundwater samples (n = 1 to 38 for all groundwater wells (*GW*, gray) and n = 1 to 11 for the riparian wells (*RP*, red), 1,000 repetitions) based on samples taken during snapshot Campaigns I (upper panels) and II (lower panels). The horizontal lines indicate the isotopic composition of baseflow at the outlet during the snapshot campaign (BF_{ss}, dashed lines) and prior to the event (BF_{pe}, solid line). Note that all groundwater wells (*GW*) also includes the riparian wells.

The range in the calculated variability of the groundwater isotopic composition also decreased with increasing sample size (Figure 5). The 5th to 95th percentiles of the standard deviation of the isotopic composition of the groundwater for six randomly selected groundwater samples were 2.7‰ and 7.7‰ for Campaign I and 1.9‰ and 5.5‰ for Campaign II. For nine randomly selected samples, they were 3.2‰ and 7.1‰ for Campaign I and 2.5‰ and 5.1‰ for Campaign II. The difference between the 5th and 95th percentiles of the calculated standard deviation of the groundwater was less than 1.2‰ (which equals twice the accuracy), as soon as more than 29 respective 22 random samples were used to determine the variability in δ^2H of groundwater.

4.3. Sensitivity of Two-Component Hydrograph Separation to the Characterization of the Pre-Event Water Composition

The pre-event water fractions (f_{pe}) were highest for Event A (range: 0.85 to 1) and lowest for Event B (range: 0.65 to 0.95) when the baseflow sample taken before the event was used to characterize the pre-event water composition (BF_{pe}, solid black lines in Figure 6, Table 4). Using different (riparian) wells to characterize C_{pe} resulted in a large range of the pre-event water fractions; the maximum difference in f_{pe} for samples from individual riparian wells ranged from 0.28 to 0.47 for Events A and Event C, respectively (gray and red lines in Figure 6 for all groundwater (*GW*) and all riparian groundwater (*RP*), respectively). The difference between the minimum f_{pe} calculated when a baseflow sample (BF_{ss} or BF_{pe}) was used to characterize C_{pe} and when the average composition of all riparian wells (RP_{avg}) was used, varied between 0.03 for Event B and 0.06 for Event A (Table 4). The difference in the minimum f_{pe} calculated using the baseflow sample

Water Resources Research 10.1029/2019WR026855

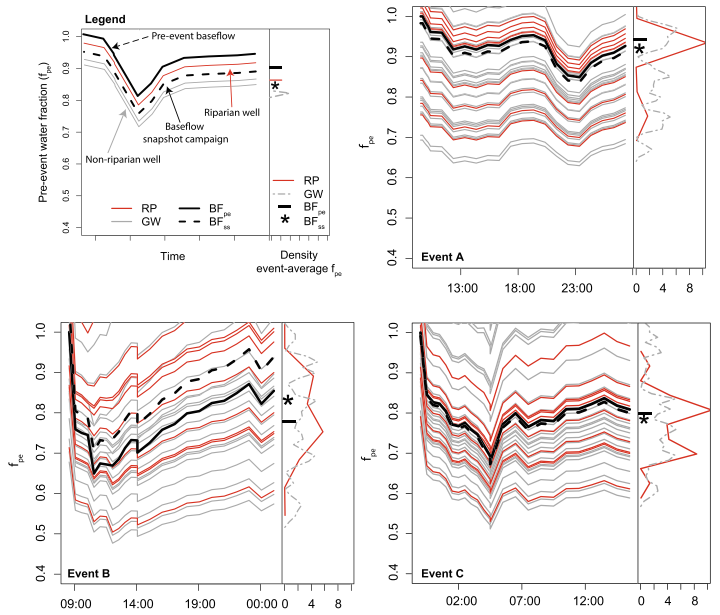

Figure 6. Time series of the calculated pre-event water fraction (f_{pe}) for Events A–C using δ^2H as a tracer and the pre-event baseflow sample (BF_{Pe}, solid black line), the baseflow sample taken during the snapshot sampling campaigns (BF_{ss}, dashed black lines), each sample from a riparian well (*RP*, red lines), and all other groundwater wells (*GW*, gray lines) to represent the isotopic composition of the pre-event water (C_{pe}), as well as the frequency distribution of the event-averaged pre-event water fraction (kernel density plot, right side of each subplot) for each method used to represent the pre-event water composition (BF_{pe}: Black dash, BF_{ss}: Black asterisk, *RP*: red solid line, *GW*: gray dashed line). See Table 2 for a detailed explanation of the different pre-event water characterization methods. Calculations for all groundwater wells (and thus also the kernel distribution) also include all riparian wells.

or the average isotopic composition of all groundwater samples (GW_{avg}) varied between 0.01 for Event A and 0.12 for Event B.

The temporal pattern of the change in the pre-event water fraction did not depend on how the pre-event water composition was characterized (Figure 6) because the same data for stream water (C_t) and rainwater (C_e) were used for all calculations. For some stream water samples (up to half of the samples, depending on which characterization for the pre-event water was used), the calculated streamflow fractions were physically impossible (fractions >1.025 or <−0.025). This was particularly the case at the beginning or end of the event when samples from wells with a very different isotopic composition than the baseflow were used.

The event-averaged pre-event water fractions were also sensitive to the choice of the sample used to characterize the pre-event water composition (density plots in Figure 6, Table 4). The spread in the event-averaged f_{pe} was, not surprisingly, largest for the ensemble of the calculations based on the individual well samples, as they spanned the whole range of possible isotopic compositions from which the average groundwater composition was calculated (*GW*, Table 3 and Figure 6). Selecting only riparian wells to characterize C_{pe} resulted

Water Resources Research 10.1029/2019WR026855

Table 4
The Range (Min-Max) and Event-Averaged Pre-event Water Fractions (f_{pe}) for Stream Water at the Catchment Outlet Calculated for Different Characterizations of the Pre-event Water Composition (See Table 2): Using the Pre-event Baseflow Sample (BF_{pe}), the Baseflow Sample From the Snapshot Campaign (BF_{ss}), the Average Groundwater Composition of All Riparian Wells (RP_{avg}) or All Groundwater Wells (GW_{avg}), or for Each Riparian Well (RP) or Each Groundwater Well (GW) Individually

C_{pe} → Event↓	BF_{pe}	BF_{ss}	RP_{avg}	GW_{avg}	RP (n = 11)	GW (n = 38)
Range f_{pe} for individual sampling times (n = 24)						
A	0.85–1	0.84–0.98	0.79–0.92	0.86–1	0.78–0.97	0.66–0.99
B	0.65–0.95	0.71–1	0.68–1	0.77–1	0.59–1	0.54–1
C	0.69–0.97	0.67–0.84	0.67–0.82	0.71–0.87	0.62–1	0.59–1
		Event-averaged f_{pe}			Event averaged (± sd)	
A	0.92	0.91	0.86	0.93	0.89 ± 0.07	0.85 ± 0.10
B	0.78	0.85	0.81	0.91	0.86 ± 0.13	0.79 ± 0.13
C	0.79	0.78	0.77	0.82	0.53 ± 0.12	0.79 ± 0.11

Note. For the event-averaged pre-event water fractions calculated based on the samples from the individual wells (*RP* and *GW*), the average ± standard deviation are given. Calculated pre-event water fractions below −0.025 or above 1.025 were excluded from the calculations.

in a higher event-averaged pre-event water fraction than either a selection from all groundwater wells or a pre-event baseflow sample (Figure 7, Table 4). However, ultimately, the latter will depend on the distribution of isotopic compositions prior to each event and might thus differ from event to event.

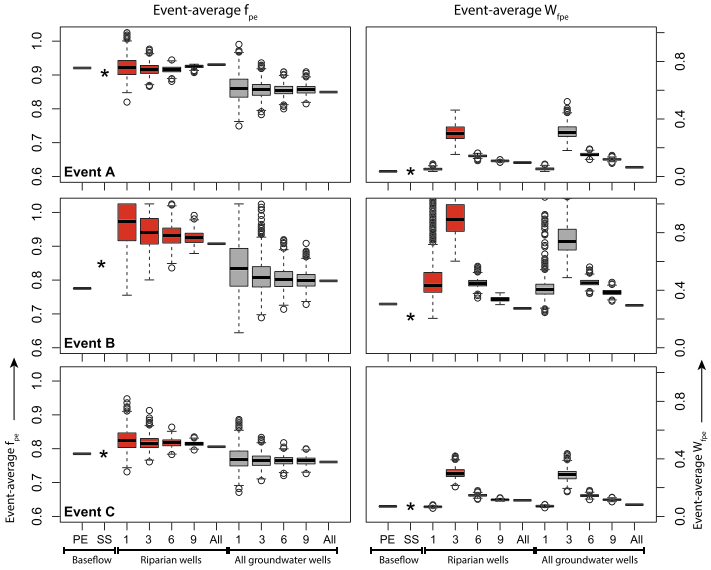

Figure 7. Boxplots of the event-averaged pre-event water fractions (f_{pe}, left) and the associated uncertainty (W_{fpe}, right) for Events A–C (rows), when the pre-event water composition is represented by a baseflow sample taken before the event (PE, dash) and a few days later during the snapshot campaign (SS, asterisk) and the average isotopic composition based on samples from one, three, six, or nine randomly selected wells in riparian areas (red) or across the entire catchment (gray) and based on the average composition of all riparian wells (all, gray, n = 11) and all wells across the catchment (all, red, n = 38). Calculations for all wells also include all riparian wells. All boxplots are based on 1,000 random selections of wells. The box extends from the 25th to the 75th percentile, the solid line represents the median, the whiskers extend to the 25th percentile −1.5*interquartile range and the 75th percentile +1.5*interquartile range; the circles represent the outliers.

Water Resources Research 10.1029/2019WR026855

Table 5
The Event-Averaged Pre-event Water Fraction (f_{pe}) and the Associated Uncertainty (W_{fpe}, 95% confidence interval, equation 4) When the Pre-event Water Composition (C_{pe}) Was Based on the Pre-event Baseflow (BF_{pe}) Sample, a Baseflow Sample Taken During the Snapshot Sampling Campaign (BF_{ss}), and the Median Event-Averaged Pre-event Water Fractions (f_{pe}) and Associated Uncertainty (W_{fpe}) When the Average Composition of Samples From One, Three, or Nine Randomly Selected Wells, or all Available Wells From the Riparian Areas (RP_1-RP_{avg}) or the Entire Catchment (GW_1–GW_{avg}) Were Used to Characterize the Pre-event Water Composition (C_{pe}) for Events A–C

Event	A		B		C	
C_{pe} characterization (n repetitions)	Event-averaged $f_{pe} \pm W_{fpe}$	10% - 90% Range f_{pe}	Event-averaged $f_{pe} \pm W_{fpe}$	10% - 90% Range f_{pe}	Event-averaged $f_{pe} \pm W_{fpe}$	10% - 90% Range f_{pe}
BF_{pe} (1)	0.92 ± 0.04^{abcd}		0.78 ± 0.30^{cdefg}		0.79 ± 0.07^{abcde}	
BF_{ss} (1)	0.91 ± 0.04^{abcd}		0.85 ± 0.22^{abcdefg}		0.78 ± 0.07^{abcde}	
RP_1 (1,000)	0.92 ± 0.05^{a}	0.88–0.97	0.96 ± 0.48^{a}	0.86–1.03	0.82 ± 0.06^{a}	0.78–0.87
RP_3 (1,000)	0.93 ± 0.30^{b}	0.89–0.94	0.94 ± 0.92^{b}	0.87–1.01	0.82 ± 0.30^{b}	0.80–0.84
RP_6 (1,000)	0.92 ± 0.14^{b}	0.89–0.94	0.93 ± 0.50^{c}	0.89–0.97	0.82 ± 0.15^{b}	0.80–0.83
RP_9 (1,000)	0.93 ± 0.11^{a}	0.90–0.93	0.93 ± 0.38^{d}	0.90–0.95	0.81 ± 0.12^{c}	0.81–0.82
RP_{avg} (1)	0.93 ± 0.09^{abc}		0.91 ± 0.27^{abcdef}		0.81 ± 0.11^{abcde}	
GW_1 (1,000)	0.86 ± 0.05^{d}	0.82–0.91	0.90 ± 0.43^{e}	0.75–0.97	0.77 ± 0.07^{d}	0.73–0.82
GW_3 (1,000)	0.86 ± 0.31^{c}	0.83–0.89	0.84 ± 0.76^{f}	0.76–0.87	0.76 ± 0.29^{e}	0.74–0.79
GW_6 (1,000)	0.86 ± 0.15^{c}	0.84–0.88	0.81 ± 0.45^{g}	0.77–0.85	0.76 ± 0.14^{e}	0.75–0.78
GW_9 (1,000)	0.86 ± 0.12^{c}	0.84–0.88	0.80 ± 0.38^{g}	0.77–0.83	0.76 ± 0.12^{e}	0.75–0.78
GW_{avg} (1)	0.86 ± 0.06^{acd}		0.78 ± 0.30^{bcdefg}		0.76 ± 0.08^{abcde}	

Note. The number of repetitions is indicated in parentheses. Calculated pre-event water fractions below −0.025 or above 1.025 were set to −0.025 or 1.025, respectively. Event-averaged f_{pe} values with different superscript letters (a–g) for an event are significantly different.

4.4. Uncertainty of the Event-Averaged Pre-Event Water Fraction

The median uncertainty in the event-averaged pre-event water fraction (W_{fpe} in equation 4 and Table 5) ranged from a low of 0.04 when using pre-event baseflow to characterize the pre-event water composition for Event A (BF_{pe}) to a high of 0.92 (median value for all combinations) when using three riparian wells to characterize the pre-event water composition (RP_3) for Event B. Overall, the calculated uncertainties in f_{pe} were smaller for Events A and C (range: 0.04 to 0.31) than for Event B (range: 0.22 to 0.92) because of the smaller variation in event water isotopic composition (standard deviation of C_e: 5.2‰ for Event A and 6.5‰ for Event C vs. 12.6‰ for Event B).

Increasing the number of samples to determine C_{pe} reduced the variability in the event-averaged pre-event water fraction and the uncertainty in the pre-event water fraction (Figure 7). W_{fpe} was largest when three samples were used to calculate the pre-event water composition due to the high t value for small sample sizes and the high standard deviation for some of the combinations of samples (see right column of Figure 5). As a result, the reduction in the median uncertainty was largest when the number of samples increased from three to six (Table 5).

The uncertainty in the pre-event water fraction (W_{fpe}) was smallest for the calculations based on a baseflow sample or one groundwater sample because we assumed that the uncertainty of the pre-event water composition (W_{Cpe} in equation 4) was equal to the measurement precision for this situation. For the uncertainty estimation for the pre-event water fraction based on the selection of three, six, or nine groundwater samples or the average composition of all (riparian) groundwater samples, W_{Cpe} was based on the standard deviation of the selected samples and corresponding t value for small sample sizes and thus, to some extent, reflects the variability in the pre-event water composition.

4.5. Event D

In addition to the three events presented above, we also sampled a 130-mm rainfall event between 31 August and 2 September 2017 (Figure 8), which exemplifies the difference between the pre-event groundwater composition and pre-event baseflow. The event lasted 51 h and caused the discharge to increase tenfold, from 0.6 to 5.9 mm h^{-1}. We sampled the groundwater across the catchment on 24 August 2017 (map with groundwater isotopic composition in supporting information S3). The average δ^2H for samples from the riparian wells was −67.2‰; the average for the samples from all 34 groundwater wells that contained water was

Water Resources Research 10.1029/2019WR026855

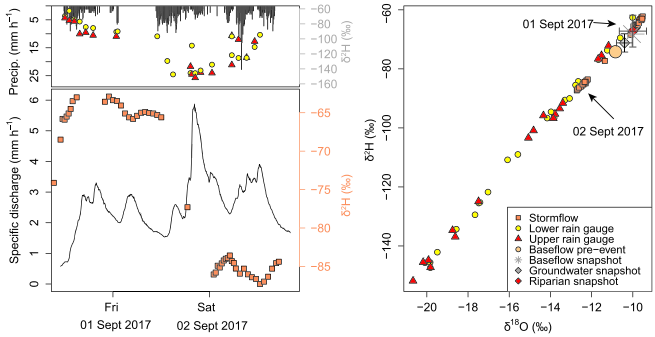

Figure 8. (left) Time series of 10-min precipitation (mm h^{-1}, bar graph) and δ^2H of rainwater sampled at the lower rain gauge (yellow circles) and upper rain gauge (red triangles) (upper plot) and specific discharge (line graph) and δ^2H of stream water (orange squares) (lower plot). (right) The dual-isotope plot (δ^2H vs $\delta^{18}O$) for the stormflow (orange squares), pre-event baseflow (orange circle), baseflow during the snapshot sampling campaign (asterisk), the average groundwater (light gray diamond) and riparian groundwater (red diamond), and the rainwater sampled at the lower (yellow circles) and upper rain gauge (red triangles). The error bars for the (riparian) groundwater samples indicate the average composition ±1 standard deviation. The rainfall samplers were full during the last part of the event, and thus, this part was not sampled.

−71.2‰. The baseflow δ^2H during the sampling campaign was −68.2‰, whereas the δ^2H of the baseflow sample taken just before the rain started on 31 August 2017 was −74.2‰. Rainfall at the start of the event was more enriched than later in the event. The average δ^2H for the first 12 rainfall samples (i.e., the rainfall sampled at two locations during the first 8 hr of the event) was −79.8‰. For the next six samples the average δ^2H was −108.3‰. The event-averaged rainfall δ^2H was −108.8‰ (range: −69.5‰ to −151.9‰).

Although baseflow and rainfall had a similar isotopic composition (−74.2‰ and −79.8‰ δ^2H, respectively), stream water became more enriched (change from −74.2‰ to −63.1‰ δ^2H, Figure 8) during the first 8 hr of the event. This suggests that neither the groundwater samples taken the week before the event nor the baseflow sample taken before the event represented the pre-event water composition that contributed to streamflow during the first hours of the event. Later (2 September), the streamflow became more depleted and reflected a mixture of baseflow/groundwater and rainfall (Figure 8).

A very simple inverse hydrograph separation calculation, assuming a pre-event water fraction of 0.79 (which is the median f_{pe} for 24 events in the neighboring Erlenbach catchment (von Freyberg et al., 2018)), suggests that the average pre-event water composition must have been approximately −61‰. However, the pre-event water fraction was likely lower (because it was a large event). Assuming a pre-event water fraction (f_{pe}) of 0.3 would imply a pre-event water composition (C_{pe}) of −43‰. These estimates of the pre-event water fraction are highly uncertain but show that the pre-event water that contributed to the streamflow had to at least be 10‰ more enriched than the baseflow sample taken before the event and also more enriched than the average composition of the groundwater measured in any of the baseflow snapshot campaigns (Kiewiet et al., 2019). In this comparison we did not consider the spatial variability in rainfall isotopic composition. However, we cannot exclude its influence on the stream water composition. It might be that part of the rainfall was more depleted (or enriched) at some locations than we sampled, so that the difference between pre-event baseflow and pre-event water might have been smaller (or larger). Daily precipitation collected at a rainfall sampler in the Erlenbach showed that there was a 17-mm rainfall event with an enriched isotopic composition (−32.4‰ δ^2H) on the evening of 24 August (i.e., between the snapshot campaign and the sampled event; see supporting information S2). For comparison, the mean isotopic composition of the daily precipitation samples taken between June and October 2017 was −48.7 ± 23.5‰ δ^2H ($n = 85$); the weighted mean composition of the precipitation was −60.7‰ δ^2H.

Water Resources Research 10.1029/2019WR026855

5. Discussion

5.1. Spatial Variability in Shallow Groundwater Composition

The snapshot groundwater sampling campaigns highlighted the large spatial variability in the shallow groundwater isotopic composition in the Studibach (standard deviations of 3.9‰ and 5.3‰ δ^2H and 0.43‰ and 0.60‰ $\delta^{18}O$ for Campaigns I and II, respectively). Large spatial variabilities in the isotopic composition of groundwater were also reported by Carey and Quinton (2005) (0.7‰ to 0.8‰ $\delta^{18}O$) and by Klaus et al. (2015) (range: 1.8‰ $\delta^{18}O$ and 8.3‰ δ^2H). However, in contrast to our observations, their samples indicated evaporative enrichment of the groundwater. Kendall et al. (2001) concluded that the variability of soil water and groundwater was ±4‰ $\delta^{18}O$. Given that a large spatial variability in the isotopic composition of groundwater is thus not uncommon, we expect that the observed variability in the Studibach is a reasonable representation of the actual spatial variability in similar small pre-alpine headwater catchments.

The basic sample size calculation to estimate the average groundwater composition with an error margin that is twice the analytical precision (1.2‰ δ^2H) suggested that for the Studibach we would need to sample 41 or 95 wells, provided that the measurements represent all hydrogeomorphic units in the catchment, as was the case in our sampling design. This number of wells and samples is unrealistic in terms of sampling effort for most catchments, even though this error margin still spans 17% to 31% of the change in the isotopic composition of stream water during the three events analyzed in this study. Results for the groundwater EC indicate that one can expect even larger uncertainties than we presented for the isotopic composition when other tracers are used for hydrograph separation (although that will admittedly depend on the tracer, the typical concentrations, and the site characteristics).

A review by Penna and van Meerveld (2019) suggested that only a third of the small catchment (<10 km^2) studies that determined the isotopic composition of groundwater sampled five or more wells. The results from this study suggest that for five randomly selected samples the 5th percentile of the standard deviation was 2.2‰ and 1.6‰ for snapshot sampling Campaigns I and II, respectively. This is already larger than the 1‰ δ^2H variation for pre-event water suggested by McDonnell et al. (1991) and indicates that a reasonable number of groundwater samples can give a rough estimate of the spatial variability in the pre-event water composition. It also indicates that we should increase the number of spatially distributed samples beyond the typical sampling effort if we want to characterize the spatial variability in the pre-event water composition and obtain a better estimate of the uncertainty of hydrograph separation results.

Although pre-event water dominated the isotopic composition of streamflow, event water also impacted its composition. The event water compositioncan vary strongly in space (Fischer et al., 2017), and this variability might have been larger than captured with our measurement set-up. Therefore, we cannot exclude that the unexpected changes in stream water composition are (partially) caused by spatial variability in the event water composition, rather than only by the spatial variability of pre-event water. For example, the sudden change in stream water isotopic composition during peakflow of Event C (Figure 4) might have been influenced by late Event B rainfall, but it could have also been caused by rainfall that fell during a short episode of increased rainfall intensity (and would thus have been more depleted) or due to spatial variability in event water. Similarly, the unexpected stream water composition for Event D could also partially be due to event water variability.

5.2. Baseflow Does Not Reflect the Catchment Average Groundwater Composition

The results from the snapshot groundwater sampling campaigns show that baseflow is not per se a mixture of all groundwater in the catchment, nor a mixture of all riparian groundwater (Figure 5, and see Kiewiet et al. (2019) for a comparison using multiple tracers). It is not surprising that the composition of baseflow does not reflect the catchment average groundwater composition. Only a small part of the Studibach is hydrologically connected to the stream during baseflow conditions (Rinderer et al., 2019) because of the steep slopes and differences in hydraulic conductivity between the different landscape elements (i.e., the forested hillslopes retain much less water than the flatter grassland sites, where the hydraulic conductivity is much lower). Hence, with the expansion of the connected contributing area during events, different landscape elements with different isotopic signatures contribute to the streamflow mixture (Figure 1).

Singh et al. (2016) report a similar observation when they sampled the isotopic composition of shallow groundwater in two adjacent headwater catchments at the Coweeta Hydrologic Laboratory (North

Water Resources Research 10.1029/2019WR026855

Carolina, USA). They found that the spatial variability of the shallow groundwater composition was larger than the variability in baseflow in both catchments. They showed that during low baseflow conditions (low connectivity) and high baseflow conditions (high connectivity), groundwater and baseflow were almost identical and baseflow was spatially least variable, while during transition periods, the spatial variability in the baseflow composition was largest. Soulsby et al. (2007) similarly showed that the sources of baseflow shifted with changes in the hydrological conditions in the Bruntland Burn catchment (Scotland). They suggested that the much higher stream water alkalinity during low flow conditions reflected the smaller influence of soil water seepage and larger influence of the well-buffered groundwater.

A better understanding of which areas are hydrologically connected to the stream can aid the development of comprehensive sampling schemes to determine the pre-event water composition. This might avoid the calculation of physically impossible (pre-)event water fractions (<0 or >1) to streamflow in hydrograph separation analyses (McDonnell et al., 1991). To accomplish this, we deem it important to also consider events for which isotope hydrograph separation does not "work," such as Event D on 31 August 2017 (Figure 8). Unfortunately, such events now often remain unpublished. Unexpected responses in stream water composition are, potentially, the result of heterogeneity in the pre-event water composition and could provide important information to test our hypotheses on runoff generation processes and evaluate interpretations of previous hydrograph separation results. The event on 31 August 2017 highlighted that at the beginning of the event stormflow was not a mixture of the sampled baseflow and rainfall and that a different type of water contributed to stormflow.

5.3. Sensitivity and Uncertainty of Hydrograph Separation Results

The calculated uncertainties for the pre-event water fractions (W_{fpe}) that we obtained (Figure 7 and Table 5) are either comparable to or larger than the uncertainties reported in other studies. Penna et al. (2017) used EC and δ^2H in a three-component hydrograph separation to quantify snowmelt fractions in streamflow for the Rio Vauz catchment (Italy) and used the standard deviation of stream water δ^2H and samples from springs (collected over a 5-year period) to quantify the uncertainty of the composition of the pre-event water (W_{Cpe}). They used a 70% confidence interval and calculated an uncertainty range between 8% to 10% ($f_{pe} \pm 0.08$ to 0.10) for two of the catchments and 6% to 21% ($f_{pe} \pm 0.06$ to 0.21) for the catchment with the smallest snowmelt fraction. Pellerin et al. (2008) used EC to perform hydrograph separation for 19 rainfall events in the Saw Mill Brook watershed (Massachusetts, USA). They assumed ± 10 μS cm^{-1} for W_{Ce} and the measured standard deviation over 24-hr baseflow periods for W_{Cpe} (± 52 to 130 μS cm^{-1}). The reported uncertainties varied between 1% and 10% (median: 4.5%). The minimum uncertainties presented by Penna et al. (2017) and Pellerin et al. (2008) correspond quite closely to our minimum uncertainties (Table 5), but our event-averaged uncertainties were much higher (range W_{fpe}: 0.14 to 0.50 when using the spatial variability based on six riparian groundwater samples; RP_6). Although we only considered the uncertainty due to the spatial variability in the pre-event water composition, it is important to note that uncertainties in hydrograph separations due to the spatial variability in event water can be just as large. For instance, Cayuela et al. (2019) reported an uncertainty of 0.01–0.14 for the Can Vila catchment (Spain). The f_{pe} estimations of Lyon et al. (2009) and Fischer et al. (2017) differed more than 50% between computations based on different rainfall sampling locations for the Upper Sabino catchment (Arizona, USA) and Zwäckentobel (Switzerland), respectively. Altogether, these findings suggest that uncertainties in the pre-event water fractions can be large, even when the variability that is included in the calculations might still be smaller than the actual variability.

The sensitivity of hydrograph separation results to the variability of the pre-event water compositions has been addressed previously, by repeating the hydrograph separation calculations for a range of (observed or estimated) pre-event water compositions. McDonnell et al. (1991) found that a ± 1‰ δ^2H range in the pre-event water composition led to a pre-event water fraction that was within ± 5% (i.e., $f_{pe} \pm 0.05$) of the original estimate, except for the peak flow sample, for which the range was larger. They also found that shifting to three-component hydrograph separation increased uncertainties compared to a two-component approach. Carey and Quinton (2005) varied the event water composition with 2‰ δ^{18}O and found that the calculated pre-event water fraction changed up to 19% (i.e., $f_{pe} \pm 0.19$) for a three-component hydrograph separation using EC and δ^{18}O. Our results demonstrate that the event-averaged pre-event water fraction can differ by 10% to 14% (i.e., $f_{pe} \pm 0.10$ to 0.14) when using three samples from different randomly selected wells

Water Resources Research 10.1029/2019WR026855

for the pre-event water characterization and by 4% to 13% when using seven samples, even if the samples are all from riparian areas (Figure 6).

In most studies, the uncertainty of the hydrograph separation results is estimated with the Gaussian standard error method of Genereux (1998; equation 4 in this manuscript). The uncertainties of W_{Cpe} and W_{Ce} are ideally based on the observed variability in the catchment. However, that requires knowledge about the spatial and temporal variation in the pre-event and event water composition. Often, only a pre-event streamwater (or baseflow) sample is available (Penna & van Meerveld, 2019), and sometimes researchers even assume that W_{Cpe} is equal to the analytical precision of the isotope analyzer (e.g., Jefferson et al., 2015) or present results without any uncertainty estimations (Qu et al., 2017; Zhao et al., 2016). It would be better to use the standard deviation of samples taken from multiple wells, to base W_{Cpe} on repeated sampling along the stream (cf. James & Roulet, 2009 or Singh et al., 2016), or to use literature values for the variability in the pre-event water composition. Using the analytical precision for W_{Cpe} due to a lack of data on the spatial variability in either the baseflow or the groundwater composition leads to an underestimation of the actual uncertainty because it neglects the variability of the pre-event water composition. This can in turn lead to a wrong interpretation of hydrological processes based on these results.

5.4. The Way Forward to Characterize the Pre-Event Water Composition for Hydrograph Separation

We have shown that a large number of samples might be needed to estimate the average pre-event water composition but that a smaller number of samples already gives an estimate of the variability (Figure 5). We also demonstrated that the outcomes of the hydrograph separation and the uncertainty estimates are sensitive to which samples and how many samples are used to characterize the pre-event water composition (Figures 6 and 7). Additionally, the data from the 130-mm event (Figure 8) and a theoretical example (supporting information S1) show that unidentified water sources can have a large effect on the calculated pre-event water fractions.

At first sight, these results might seem discouraging for hydrograph separation analyses or characterization of the pre-event water composition. One might decide to refrain from it altogether or take it as a challenge to determine a time-variable pre-event water composition that reflects the changes in the contributing areas and use this in hydrograph separation calculations (as it is done for the event water composition and is suggested by Harris et al. (1995)). However, this requires knowledge on the contributing areas and is in conflict with the simplicity and thus attractiveness of the hydrograph separation method. Instead, we are convinced that despite these challenges, isotope hydrograph separation can remain a useful toolbut that the results need to be interpreted with care. Quantifying the uncertainty and sensitivity of the analyses by also considering the spatial variability in the (pre-) event water composition is a first step toward improving our interpretations. The results of this study show that when the variability in the pre-event water composition is included, the uncertainty in the pre-event water fractions is likely larger than reported for most studies in small headwater catchments. This should be acknowledged when comparing results for different events or different catchments.

A lack of data is a challenge for any method that estimates the error in mixing fractions. We, therefore, encourage researchers to sample baseflow or groundwater at more locations than is typically the case. Our results suggest that after sampling a few wells, we would have known that the spatial variability is large and would have had a rough estimate of the variability (Figure 5). We do not think that a network with more than 30 groundwater wells is feasible (or needed) for all research areas but encourage additional studies on the spatial variability in shallow groundwater in different climatic and geologic settings so that literature values on the typical spatial variability in the shallow groundwater composition become available and can be used in other studies. In future studies, it will be essential to consider that although samples from some wells might seem uninformative for hydrograph separation (because they come from areas that might not be hydrologically connected to the stream during small or intermediately sized events), they are still essential to characterize the variability in the catchment (average) pre-event water composition that contributes to streamflow during extreme events.

AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

10.1029/2019WR026855

6. Conclusions

Isotope hydrograph separation is a powerful tool to investigate runoff sources and catchment functioning. For undisturbed headwater catchments in temperate climates, results usually show that groundwater makes up the largest portion of streamflow. However, the assumption of a constant pre-event water or groundwater composition during the event is likely violated in most of these studies because there is not a single well-mixed groundwater source. We assessed the spatial variability in the isotopic composition of groundwater in a small steep, humid headwater catchment and found that the spatial variability in varied isotopic composition of shallow groundwater was large (standard deviation: 3.9‰ and 5.3‰ δ^2H). A rough sample size estimation suggests that more than 12 wells need to be sampled to estimate the average groundwater composition within 2.5‰ δ^2H (half of the variability in the streamflow during events). The difference between the isotopic composition of baseflow and the average (riparian) groundwater ranged from 0.5‰ to 2.2‰ δ^2H. As such, the baseflow sample might represent the average pre-event water that contributes to streamflow during the event, but it might also be different because other sources (e.g., hillslopes) may only contribute to streamflow after the expansion of the contributing area. In other words, an apparent (or even physically impossible) event water contribution might be the result of a temporally varying composition of the pre-event water that contributes to streamflow.

We quantified the sensitivity of hydrograph separation results to different characterizations of the pre-event water composition by repeating the calculations for different sets of baseflow and groundwater samples. We found that hydrograph separation results based on riparian groundwater samples or a baseflow sample resulted in different calculated pre-event water fractions than when the catchment average groundwater composition was used to characterize the pre-event water composition. Even if we selected three riparian groundwater samples to characterize the pre-event water composition, the event-averaged pre-event water fractions varied by 0.07 to 0.17.

The uncertainties in the pre-event water fractions (W_{fpe}) were lowest when one baseflow sample was used to represent the pre-event water composition, but we argue that this gives a false sense of accuracy because it neglects the spatial variability in the groundwater isotopic compositions. Furthermore, we show that this sample may not represent the actual pre-event water composition that contributes to streamflow during the event. The reduction in the event-averaged uncertainty of the pre-event water composition was largest when the number of groundwater samples increased from three to six.

To summarize, our results highlight the importance of representing the variability in the pre-event water composition when applying hydrograph separation analyses to assess runoff processes. This can be achieved by, for instance, increasing the number of sampling locations or by using ranges reported in literature.

Acknowledgments

This research project would not have been possible without the help and support of many people in the lab and field. We particularly thank Barbara Herbstritt for the isotope analyses; Michael Rinderer and Benjamin Fischer for the helpful discussions and for the installation of the wells and stream gauges; the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL); Oberallmeindkorporation Schwyz (OAK), the municipality of Alpthal; and the Department of Environment of the Canton of Schwyz for the excellent cooperation. We thank the Editor, Associate Editor, and five anonymous reviewers for their helpful comments on previous versions of this manuscript. The data used in this analysis are available through the data repository of WSL (www.envidat.ch).

References

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, *24*(1), 43–69. https://doi.org/10.1080/02626667909491834

Burns, D. A., Plummer, L. N., McDonnell, J. J., Busenberg, E., Casile, G. C., Kendall, C., et al. (2003). The geochemical evolution of riparian ground water in a forested Piedmont catchment. *Groundwater*, *41*(7), 913–925. https://doi.org/10.1111/j.1745-6584.2003.tb02434.x

Buttle, M. (1994). Isotope hydrograph separations and rapid delivery of pre-event water from drainage basins. *Progress in Physical Geography*, *18*, 16–41. https://doi.org/10.1177/030913339401800102

Carey, S. K., & Quinton, W. L. (2005). Evaluating runoff generation during summer using hydrometric, stable isotope and hydrochemical methods in a discontinuous permafrost alpine catchment. *Hydrological Processes*, *19*(1), 95–114. https://doi.org/10.1002/hyp.5764

Cayuela, C., Latron, J., Geris, J., & Llorens, P. (2019). Spatio-temporal variability of the isotopic input signal in a partly forested catchment: Implications for hydrograph separation. *Hydrological Processes*, *33*, 36–46. https://doi.org/10.1002/hyp.13309

Feyen, H., Wunderli, H., Wydler, H., & Papritz, A. (1999). A tracer experiment to study flow paths of water in a forest soil. *Journal of Hydrology*, *225*(3–4), 155–167. https://doi.org/10.1016/S0022-1694(99)00159-6

Fischer, B., Aemisegger, F., Graf, P., Sodemann, H., & Seibert, J. (2019). Assessing the sampling quality of a low-tech low-budget volume-based rainfall sampler for stable isotope analysis. *Frontiers in Earth Science*, *7*, 1–8. https://doi.org/10.3389/feart.2019.00244

Fischer, B. M. C., Stahli, M., & Seibert, J. (2016). Pre-event water contributions to runoff events of different magnitude in pre-alpine headwaters. *Hydrology Research*, *48*, 28–47. https://doi.org/10.2166/nh.2016.176

Fischer, B. M. C., van Meerveld, H. J. I., & Seibert, J. (2017). Spatial variability in the isotopic composition of rainfall in a small headwater catchment and its effect on hydrograph separation. *Journal of Hydrology*, *547*, 755–769. https://doi.org/10.1016/j.jhydrol.2017.01.045

Genereux, D. (1998). Quantifying uncertainty in tracer-based hydrograph separations. *Water Resources Research*, *34*(4), 915–919. https://doi.org/10.1029/98WR00010

Hagedorn, F., Schleppi, P., Waldner, P., & Flühler, H. (2000). Export of dissolved organic carbon and nitrogen from Gleysol dominated catchments—The significance of water flow paths. *Biogeochemistry*, *50*, 137–161. https://doi.org/10.1023/A:1006398105953

Water Resources Research 10.1029/2019WR026855

Harris, D. M., McDonnell, J. J., & Rodhe, A. (1995). Hydrograph separation using continuous open system isotope mixing. *Water Resources Research*, *31*(1), 157–171. https://doi.org/10.1029/94WR01966
Hooper, R. P. (2001). Applying the scientific method to small catchment studies: A review of the Panola Mountain experience. *Hydrological Processes*, *2050*, 2039–2050. https://doi.org/10.1002/hyp.255
Inamdar, S. P., & Mitchell, M. J. (2007). Contributions of riparian and hillslope waters to storm runoff across multiple catchments and storm events in a glaciated forested watershed. *Journal of Hydrology*, *341*, 116–130. https://doi.org/10.1016/j.jhydrol.2007.05.007
Inserillo, E. A., Green, M. B., Shanley, J. B., & Boyer, J. N. (2017). Comparing catchment hydrologic response to a regional storm using specific conductivity sensors. *Hydrological Processes*, *31*(5), 1074–1085. https://doi.org/10.1002/hyp.11091
James, A. L., & Roulet, N. T. (2009). Antecedent moisture conditions and catchment morphology as controls on spatial patterns of runoff generation in small forest catchments. *Journal of Hydrology*, *377*(3–4), 351–366. https://doi.org/10.1016/j.jhydrol.2009.08.039
Jefferson, A. J., Bell, C. D., Clinton, S. M., & McMillan, S. K. (2015). Application of isotope hydrograph separation to understand contributions of stormwater control measures to urban headwater streams. *Hydrological Processes*, *29*, 5290–5306. https://doi.org/10.1002/hyp.10680
Jencso, K. G., & McGlynn, B. L. (2011). Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation. *Water Resources Research*, *47*, W11527. https://doi.org/10.1029/2011WR010666
Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Bencala, K. E., & Wondzell, S. M. (2010). Hillslope hydrologic connectivity controls riparian groundwater turnover: Implications of catchment structure for riparian buffering and stream water sources. *Water Resources Research*, *46*, W10524. https://doi.org/10.1029/2009WR008818
Jones, J. P., Sudicky, E. A., Brookfield, A. E., & Park, Y.-J. (2006). An assessment of the tracer-based approach to quantifying groundwater contributions to streamflow. *Water Resources Research*, *42*, W02407. https://doi.org/10.1029/2005WR004130
Kendall, C., McDonnell, J. J., & Gu, W. (2001). A look inside "black box" hydrograph separation models: A study at the hydrohill catchment. *Hydrological Processes*, *15*(10), 1877–1902. https://doi.org/10.1002/hyp.245
Kennedy, V. C., Zellweger, G. W., & Avanzino, R. J. (1979). Variation of rain chemistry during storms at two sites in northern California. *Water Resources Research*, *15*(3), 687–702. https://doi.org/10.1029/WR015i003p00687
Kiewiet, L., von Freyberg, J., & van Meerveld, H. J. I. (2019). Spatiotemporal variability in hydrochemistry of shallow groundwater in a small pre-alpine catchment: The importance of landscape elements. *Hydrological Processes*, 1–21. https://doi.org/10.1002/hyp.13517
Klaus, J., & McDonnell, J. J. (2013). Hydrograph separation using stable isotopes: Review and evaluation. *Journal of Hydrology*, *505*, 47–64. https://doi.org/10.1016/j.jhydrol.2013.09.006
Klaus, J., McDonnell, J. J., Jackson, C. R., Du, E., & Griffiths, N. A. (2015). Where does streamwater come from in low-relief forested watersheds? A dual-isotope approach. *Hydrology and Earth System Sciences*, *19*(1), 125–135. https://doi.org/10.5194/hess-19-125-2015
Landwehr, J. M., & Coplen, T. B. (2006). Line-conditioned excess: A new method for characterizing stable hydrogen and oxygen isotope ratios in hydrologic systems (IAEA-CSP--26/P). International Atomic Energy Agency (IAEA).
Laudon, H., Hemond,, H. F., Krouse, R., & Bishop, K. H. (2002). Oxygen 18 fractionation during snowmelt: Implications for spring flood hydrograph separation. *Water Resources Research*, *38*(11), 1258. https://doi.org/10.1029/2002WR001510
Lyon, S. W., Desilets, S. L. E., & Troch, P. A. (2009). A tale of two isotopes: Differences in hydrograph separation for a runoff event when using δD versus $\delta^{18}O$. *Hydrological Processes*, *23*, 2095–2101. https://doi.org/10.1002/hyp.7326
McCallum, J. L., Cook, P. G., Brunner, P., & Berhane, D. (2010). Solute dynamics during bank storage flows and implications for chemical base flow separation. *Water Resources Research*, *46*, W07541. https://doi.org/10.1029/2009WR008539
McDonnell, J., Stewart, M., & Owens, I. (1991). Effect of catchment-scale subsurface mixing on stream isotopic response. *Water Resources Research*, *27*(12), 3065–3073.
McDonnell, J. J., Bonell, M., Stewart, M. K., & Pearce, A. J. (1990). Deuterium variations in storm rainfall: Implications for stream hydrograph separation. *Water Resources Research*, *26*(3), 455–458. https://doi.org/10.1029/WR026i003p00455
McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., et al. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research*, *43*, W07301. https://doi.org/10.1029/2006WR005467
McGlynn, B. L., & McDonnell, J. J. (2003). Quantifying the relative contributions of riparian and hillslope zones to catchment runoff. *Water Resources Research*, *39*(11), 1310. https://doi.org/10.1029/2003wr002091
Mohn, J., Schürmann, A., Hagedorn, F., Schleppi, P., & Bachofen, R. (2000). Increased rates of denitrification in nitrogen-treated forest soils. *Forest Ecology and Management*, *137*(1–3), 113–119. https://doi.org/10.1016/S0378-1127(99)00320-5
Molnar, P., Densmore, A. L., McArdell, B. W., Turowski, J. M., & Burlando, P. (2010). Analysis of changes in the step-pool morphology and channel profile of a steep mountain stream following a large flood. *Geomorphology*, *124*(1–2), 85–94. https://doi.org/10.1016/j.geomorph.2010.08.014
Pearson, E. S., & Hartley, H. O. (1954). *Biometrika tables for statisticians*. Cambridge, England: Cambridge University Press.
Pellerin, B. A., Wollheim, W. M., Feng, X., & Charles, J. V. (2008). The application of electrical conductivity as a tracer for hydrograph separation in urban catchments. *Hydrological Processes*, *22*, 1810–1818. https://doi.org/10.1002/hyp
Penna, D., & van Meerveld, H. J. (2019). Spatial variability in the isotopic composition of water in small catchments and its effect on hydrograph separation. *WIREs Water*, 1–33. https://doi.org/10.1002/wat2.1367
Penna, D., van Meerveld, H. J., Oliviero, O., Zuecco, G., Assendelft, R. S., Dalla Fontana, G., & Borga, M. (2015). Seasonal changes in runoff generation in a small forested mountain catchment. *Hydrological Processes*, *29*(8), 2027–2042. https://doi.org/10.1002/hyp.10347
Penna, D., van Meerveld, H. J., Zuecco, G., Dalla Fontana, G., & Borga, M. (2016). Hydrological response of an Alpine catchment to rainfall and snowmelt events. *Journal of Hydrology*, *537*, 382–397. https://doi.org/10.1016/j.jhydrol.2016.03.040
Penna, D., Zuecco, G., Crema, S., Trevisani, S., Cavalli, M., Pianezzola, L., et al. (2017). Response time and water origin in a steep nested catchment in the Italian Dolomites. *Hydrological Processes*, *31*(4), 768–782. https://doi.org/10.1002/hyp.11050
Qu, S., Wang, Y., Zhou, M., Liu, H., Shi, P., & Yu, Z. (2017). Temporal O and deuterium variations in hydrologic components of a small watershed during a typhoon event. *Isotopes in Environmental and Health Studies*, *53*(2), 152–153. https://doi.org/10.1080/10256016.2016.1205590
Rinderer, M., Van Meerveld, H. J., & McGlynn, B. L. (2019). From points to patterns: Using groundwater time series clustering to investigate subsurface hydrological connectivity and runoff source area dynamics. *Water Resources Research*, *55*, 5784–5806. https://doi.org/10.1029/2018WR023886
Rinderer, M., van Meerveld, H. J., & Seibert, J. (2014). Topographic controls ond shallow groundwater levels in a steep, prealpine catchment. *Water Resources Research*, *50*, 6067–6080. https://doi.org/10.1002/2013WR015009
Rodhe, A. (1981). Spring flood: Meltwater or groundwater? *Nordic Hydrology*, *12*, 21–30.

Rodhe, A. (1987). *The origin of streamwater traced by oxygen-18*, *Report series A N° 41*, Uppsala (S): Uppsala University, Dept Phys. Geogr., Div. Hydrol.

Rücker, A., Boss, S., Kirchner, J. W., & von Freyberg, J. (2019). Monitoring snowpack outflow volumes and their isotopic composition to better understand streamflow generation during rain-on-snow events. *Hydrology and Earth System Sciences*, *23*, 2983–3005. https://doi.org/10.5194/hess-23-2983-2019

Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*, (3rd: ed.). New York: Chapman and Hall/CRC. https://doi.org/10.1201/9781420036268

Singh, N. K., Emanuel, R. E., & McGlynn, B. L. (2016). Variability in isotopic composition of base flow in two headwater streams of the southern Appalachians. *Water Resources Research*, *52*, 4264–4279. https://doi.org/10.1002/2015WR018463

Sklash, M. G., & Farvolden, R. N. (1979). The role of groundwater in storm runoff. *Journal of Hydrology*, *43*, 45–65.

Soulsby, C., Tetzlaff, D., van den Bedem, N., Malcolm, I. A., Bacon, P. J., & Youngson, A. F. (2007). Inferring groundwater influences on surface water in montane catchments from hydrochemical surveys of springs and streamwaters. *Journal of Hydrology*, *333*(2–4), 199–213. https://doi.org/10.1016/j.jhydrol.2006.08.016

Tetzlaff, D., Birkel, C., Dick, J., Geris, J., & Soulsby, C. (2014). Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions. *Water Resources Research*, *50*, 969–985. https://doi.org/10.1002/2013WR014147

Uhlenbrook, S., & Hoeg, S. (2003). Quantifying uncertainties in tracer-based hydrograph separations: a case study for two- , three- and five-component hydrograph separations in a mountainous catchment. *453*, 431–453. https://doi.org/10.1002/hyp.1134

van Meerveld, H. J. I., Fischer, B. M. C., Rinderer, M., Stähli, M., & Seibert, J. (2018). Runoff generation in a pre-alpine catchment: A discussion between a tracer and a shallow groundwater hydrologist. *Cuadernos de Investigación Geográfica*, *44*(2), 429–452. https://doi.org/10.18172/cig.3349

von Freyberg, J., Studer, B., Rinderer, M., & Kirchner, J. W. (2018). Studying catchment storm response using event- and pre-event-water volumes as fractions of precipitation rather than discharge. *Hydrology and Earth System Sciences*, *22*(11), 5847–5865. https://doi.org/10.5194/hess-22-5847-2018

Zhao, P., Tang, X., Zhao, P., Zhang, W., & Tang, J. (2016). Mixing of event and pre-event water in a shallow Entisol in sloping farmland based on isotopic and hydrometric measurements, SW China. *Hydrological Processes*, *30*, 3478–3493. https://doi.org/10.1002/hyp.10867

Zuecco, G., Penna, D., Borga, M., & van Meerveld, H. J. (2016). A versatile index to characterize hysteresis between hydrological variables at the runoff event timescale. *Hydrological Processe*, *30*, 1449–1466. https://doi.org/10.1002/hyp.10681